Inventory & Monitoring Program
National Park Service
U.S. Department of the Interior
Great Lakes Network

Data Management Plan
Great Lakes Inventory & Monitoring Network

National Park Service
2800 Lake Shore Dr. E, Suite D
Ashland, WI 54806

Mark Hart
Network Data Manager

&

Ulf Gafvert
Network GIS Specialist

Version: September 2006

File Name: GLKN_DataMgmtPlan_20060928.doc

Recommended Citation:
Hart, M., and U. Gafvert. Editors, 2006. Data management plan: Great Lakes Inventory and Monitoring Network. National Park Service Great Lakes Inventory and Monitoring Network Report. GLKN/2006/20.

Topic(s):
Data management, Planning, Administrative, Interdisciplinary

Theme Keywords:
Reports, protocols, guidelines, plans

Placename Keywords:
Wisconsin, Minnesota, Michigan, Indiana

Initial Distribution:
Website: *www.nature.nps.gov/im/units/GLKN/DataMgt.htm*

Data Management Plan
Great Lakes Inventory and Monitoring Network

Midwest Region Approval Signatures

_____ _____
Carmen Thomson, Regional Coordinator Date

Great Lakes Network Approval Signatures

_____ _____
Bill Route, GLKN CoordinatorDate

_____ _____
Mark Hart, GLKN Data Manager Date

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

Information is the common currency among the activities and staff involved in the stewardship of natural resources for the National Park Service (NPS). This Executive Summary describes the Great Lake Inventory and Monitoring Network's (GLKN) data management strategy. The Data Management Plan (DMP) is a guide for current and future project managers and GLKN staff to ensure the continuity and documentation of data management procedures over time. The DMP also refers to other guidance documents and standard operating procedures which convey the specific standards and steps for achieving the Network's data management goals.

The Data Management Plan focuses on the processes used to:

- acquire, store, manage and archive data,
- ensure data quality,
- document and disseminate data, and
- ensure the long-term access to and utility of data.

1. Data Management Goals

The overall goal of the NPS Inventory and Monitoring (I&M) Program is to provide scientifically and statistically sound data to support management decisions for the protection of park resources. The data management mission of the I&M Program is to ensure the quality, interpretability, security, longevity and availability of our natural resource data.

The goals of the GLKN Data Management Plan are to:

- *Goal 1* - Ensure the high quality and long-term availability of the ecological data and related analyses produced from the Network's inventory and monitoring work.
- *Goal 2* - Initiate and invest in data management activities based on data and information needs defined in Network monitoring protocols and inventory study plans.
- *Goal 3* - Integrate data management activities with all aspects and at all stages of Network business.
- *Goal 4* - Specify data stewardship responsibilities for all personnel.
- *Goal 5* - Collaborate internally and externally to address data management issues with individuals at all organizational levels.

2. Priorities of Natural Resource Data

The priorities for Network data management efforts are to:

- Produce and curate high-quality, well-documented data originating within the I&M Program
- Assist with data management for current projects, legacy data and data originating outside the I&M Program that complement program objectives
- Help ensure good data management practices for park-based natural resource projects that are just beginning to be developed and implemented

3. Data Stewardship Roles and Responsibilities

Every individual involved in the I&M Program has data stewardship responsibilities, whether they are involved in the production, analysis, management, or end use of data. It is critical that individuals perform their duties in accordance with the Data Management Plan and the specific Vital Signs monitoring protocols. Monitoring protocols will describe specific roles and responsibilities in detail. Senior Network staff (Chapter 8, Route and Elias 2005) are responsible for ensuring that data management procedures are followed.

4. Data and Information Workflow

Understanding the data life cycle, or flow of data throughout a project, will help identify and manage the resources needed to produce and provide quality data. For data management to be effective, it must occur throughout all stages of the project.

There are six distinct stages to each project:

1) Project initiation
2) Planning and approval
3) Design and testing
4) Implementation
5) Product integration
6) Evaluation and closure

The GLKN will implement a project-tracking database to monitor the flow of information for each GLKN project. Most notably, this database tracks project status, changes to protocols, and archiving and distribution of deliverables.

5. Infrastructure and System Architecture

Infrastructure refers to the network of computers and servers that our information systems are built upon. The GLKN relies heavily on national, regional, and GLKN information technology (IT) personnel and resources to maintain its computer infrastructure. This includes, but is not limited to: configuration of computers, servers and other related hardware, software installation and support, email administration, security updates, virus-protection, telecommunications, computer networking, and backups of servers.

The infrastructure supports these required functions:

- Provides a central repository for master datasets
- Provides controlled subsets of data for local computing
- Provides a means for uploading and downloading data for both NPS and public
- Supports desktop and internet applications
- Provides security, stability, and backups

The GLKN has developed procedures to maintain, store and archive data to ensure that data and related documents (digital and analog) are both accessible and secure. Content, format, and documentation must be up-to-date so that the data can be easily accessed and properly used. Data must also be physically secure against environmental hazards, catastrophe, and human malice.

Most data maintenance will be performed on the GLKN file server and on Service-wide servers maintained by the I&M Program. The GLKN administrative, IT, and data management staff are responsible for automated daily and weekly data backups. Data and information on GLKN and NPS servers will be kept current and all updates will be described in accompanying documentation. Information files will be properly cataloged and maintained on the GLKN website, and the latest versions of primary data will be available in formats that reflect common usages in the resource management community.

Project data will be electronically archived as stand-alone products and will include:

- Project documentation
- Data in raw, verified, and analyzed conditions
- Respective metadata
- Supporting files (e.g., digital photographs and maps)
- All associated reports

Final deliverables from project data will be added to existing libraries and databases.

6. Database Design Strategies

The project manager and the data manager will work together to develop conceptual and logical data models to understand the data life cycle and flow of the data collection process. It is necessary to understand where data collection begins (for example, a visit to a site) and what steps are involved in data processing. Understanding conceptual and logical relationships can also help identify how the project information can best be presented.

Understanding the relationships between data components is the key to successfully developing and using a database. If the relationships are misunderstood, data entry may be tedious and data output may be cumbersome.

The GLKN Data Management Plan specifies the standards by which data will be handled. Data management elements, or principles common to more than one Vital Sign, will be managed in a manner that enhances data integrity and allows for comparison of data across the Network.

7. Acquiring and Processing Data

The types of data handled by the I&M Program fall into three general categories. *Program data* are produced by projects that are either initiated (funded) by the I&M Program or involve the I&M Program in another manner (e.g., natural resource inventories and Vital Signs monitoring projects). *Non-program legacy (existing) data* are produced by NPS entities without the involvement of the I&M Program (e.g., park or regional projects). *Non-program external data* are produced by agencies or institutions other than the National Park Service (e.g., weather and air quality data).

Most data acquired by the Network will be collected during field-based inventory and monitoring studies or will be discovered through data mining initiatives. The methods and tools required for the collection of field data (e.g., paper data forms, field computers, automated data loggers, and GPS units) will be specified in individual monitoring protocols and study plans. Field crew members must closely follow the standard

operating procedures (SOPs) in the project protocol. Techniques for handling data acquired from non-program sources, such as data downloaded from other agencies, will also be specified in individual monitoring protocols.

8. Ensuring Data Quality

High quality data and information are vital to the credibility and success of the I&M Program, and everyone plays a part in ensuring that products conform to data quality standards.

Specific procedures to ensure data quality must be included in the protocols for each Vital Sign. Although many quality assurance/quality control (QA/QC) procedures depend on the individual Vital Signs being monitored, some general concepts apply to all.

Examples of QA/QC practices include:

- Training field crew members
- Standardizing field data sheets with descriptive data dictionaries
- Use of handheld computers and data loggers
- Maintenance and calibration of equipment
- Procedures for handling data in the field
- Incorporating database features to minimize transcription errors (e.g., validation rules, range limits, pick lists, routines to import data from data loggers)
- Verification and validation, including automated error-checking database routines

Quality assurance methods should be in place at the inception of any project and should continue through all stages of the project. The final step in project quality assurance is the preparation of summary documentation that assesses the overall quality of the data. The project manager will compose a statement of data quality that will be incorporated into the formal metadata. Metadata for each dataset will also include information on quality assurance procedures specific to the project and results of the review.

9. Data Documentation

Documenting datasets, data sources, and the methodology by which the data were acquired establishes the basis for interpreting the data and using it appropriately. At a minimum, all data managed by the Network will require documentation of the project, formal metadata compliant with Federal Geographic Data Committee (FGDC) standards, and data dictionaries and Entity Relationship Diagrams (ERDs) for all tabular databases.

Data documentation will be available via the GLKN website as well as the national I&M Program's NR-GIS Data Store.

10. Data Analysis and Reporting

The mission of the I&M Program is to provide useful information to managers and scientists; therefore, providing meaningful results from data summary and analysis is a cornerstone of the GLKN data management program. Each monitoring protocol establishes requirements for scheduled and requested data analysis and reporting. Based on such requirements, the associated databases for the protocols will include functions to summarize and report directly from the database and will allow output in formats that can

be easily imported to other analysis software programs. In addition to tabular and charted summaries, the Network provides maps of natural resource data and GIS analysis products to communicate spatial locations, relationships and geospatial model results. In Chapter 7 of the GLKN Monitoring Plan, Route and Elias (2005) detail the Network's analysis and reporting schedule and procedures.

11. Data Dissemination

The GLKN data dissemination strategy aims to ensure that:

- Data are easily discoverable and obtainable
- Only data subjected to complete quality control are released, unless release is necessary in response to a Freedom of Information Act (FOIA) request
- Distributed data are accompanied by appropriate documentation
- Sensitive data are identified and protected from unauthorized access and inappropriate use

Users have various means at their disposal to browse, search, and acquire Network data and supporting documents. Data products can be accessed via:

- Links to public data products on a GLKN public website. Data may be accessed using spatial and/or tabular queries and presentations
- The NR-GIS Data Store. Distribution instructions for each dataset will be provided in the respective metadata
- Service-wide databases, such as NPSTORET, NPSpecies, and NatureBib
- Regional, Network, or park data servers protected with read-only access
- External repositories such as the EPA STORET, U.S. Geological Survey, U.S. Forest Service, and the Midwestern Regional Climatic Center
- FTP sites, CDs, DVDs, or hard drives, as appropriate

12. Ownership, FOIA, and Sensitive Data

Great Lakes Network products are considered property of the NPS; however, the Freedom of Information Act (FOIA) establishes that any person may access federal agency records that are not protected from disclosure by exemption or by special law enforcement record exclusions. The NPS is directed to protect information about the nature and location of sensitive park resources under one Executive Order and four resource confidentiality laws:

- Executive Order No. 13007: Indian Sacred Sites
- National Parks Omnibus Management Act (NPOMA; 16 U.S.C. 5937)
- National Historic Preservation Act (16 U.S.C. 470w-3)
- Federal Cave Resources Protection Act (16 U.S.C. 4304)
- Archaeological Resources Protection Act (16 U.S.C. 470hh)

When any of these regulations are applicable, public access to data can be restricted. If disclosure could result in harm to natural resources, the records may be classified as 'protected' or 'sensitive'. The NPS recognizes the following resources as sensitive:

- Endangered, threatened, rare, or commercially valuable National Park System resources
- Mineral or paleontological sites
- Objects of cultural patrimony
- Significant caves

The GLKN will comply with all FOIA restrictions regarding the release of data and information, as instructed in NPS Director's Order #66 and accompanying Reference Manuals 66A and 66B (currently in development). Managing natural resource information that is sensitive or protected requires the following steps:

- Identification of potentially sensitive resources
- Compilation of all records relating to those resources
- Determination of which data must not be released in a public forum
- Management and archiving of those records to avoid their unintentional release

Classification of sensitive data will be the responsibility of Network staff, park superintendents (or their delegates), and project managers. Network staff will classify sensitive data on a case-by-case, project-by-project basis and will work closely with project managers to ensure that potentially sensitive park resources are identified, that information about these resources is tracked throughout the project, and that potentially sensitive information is removed from documents and products that will be released outside the Network.

13. Natural History Archiving and Records Management

In most instances, administrative documents, natural history specimens, photographs, audio tapes and other materials are essential companions to the digital data. Direction for managing many of these materials (as well as digital materials) is provided in NPS Director's Order 19: Records Management (2001) and its appendix, NPS Records Disposition Schedule (NPS-19 Appendix B, revised 5-2003). Director's Order 19 states that all records of natural and cultural resources and their management are considered mission-critical records (necessary for fulfillment of the NPS mission) and must be permanently archived.

The GLKN data management strategy includes helping project managers comply with archival directives. Whenever possible, physical items that are products of a project (e.g., reports, maps, photographs, or notebooks) will be cataloged and archived by the park(s) involved with the project. When this is not possible, these physical items will be stored in GLKN offices. Physical specimens, such as plants and animals, will be accessioned and housed at appropriate archival institutions.

14. Water Quality Data

Water quality data are managed according to guidelines from the NPS Water Resources Division. In accordance with these guidelines, the desktop database application NPSTORET will be used to enter, store, document, and transfer water quality data. The GLKN oversees the use of NPSTORET per the Network's water quality monitoring

protocols and ensures that data are transferred at least annually to the NPS Water Resource Division for upload to the STORET database.

15. Implementation

The GLKN Data Management Plan contains practices that may be new to staff and principal investigators. With a few exceptions, however, the DMP does not include any new requirements. Almost every requirement stipulated in the Plan comes from law, Director's Orders, or the I&M Program. The DMP helps put these requirements into context and provides operational guidance to meet these requirements.

Credits

Significant portions of this executive summary were adapted from concepts and material developed by Gordon Dicus (Pacific Island Network).

1. INTRODUCTION

Collecting natural resource data is our first step toward understanding the ecosystems within our National Parks. These ecosystems are evolving, as is our knowledge of them and how they work. We use these "raw" data to analyze, synthesize, and model aspects of ecosystems. In turn, we use our results and interpretations to make decisions about the Parks' vital natural resources. Thus, *data* collected by researchers and maintained through sound data management practices will become *information* through analyses, syntheses, and modeling.

Information is the common currency among the many and various activities and people involved in stewardship projects throughout the National Park Service. These projects include park planning, creating inventories, short-term and long-term monitoring, restoration, control of invasive species and other species management, fire management, trail and road maintenance, law enforcement, and the communication of natural resource information to the public.

The Inventory and Monitoring Program (I&M; *www.nature.nps.gov/im*) represents a long-term commitment by the National Park Service (NPS) to assess and document the status and trends of park ecological resources. In 1998, the National Parks Omnibus Management Act established a framework for the I&M Program which fully integrates natural resource monitoring and other scientific activities into the management processes of the National Park System.

The Omnibus Management Act (1998) charges the Secretary of the Interior to "continually improve the ability of the National Park Service to provide state-of-the-art management, protection, and interpretation of and research on the resources of the National Park System," and to "… assure the full and proper utilization of the results of scientific studies for park management decisions." Section 5934 of the Act requires the Secretary of the Interior to develop a program of "inventory and monitoring of National Park System resources to establish baseline information and to provide information on the long-term trends in the condition of National Park System resources."

To carry out this mission, the NPS initiated a Service-wide, natural resource Inventory and Monitoring Program encompassing approximately 270 park units with significant natural resources. These park units are grouped into 32 networks based on similar geography and ecology. Each network works towards the following goals of the I&M Program:

- Establish natural resource inventory and monitoring as a standard practice throughout the National Park System that transcends traditional program, activity, and funding boundaries.
- Inventory the natural resources and park ecosystems under National Park Service stewardship to determine their nature and status.
- Monitor park ecosystems to better understand their dynamic nature and condition and to provide reference points for comparisons with other, altered environments.
- Integrate natural resource inventory and monitoring information into National Park Service planning, management, and decision making.

- Share National Park Service accomplishments and information with other natural resource organizations and form partnerships for attaining common goals and objectives.

The last two of these goals can only be achieved through the development of a modern information management infrastructure (i.e., staffing, hardware, software) and good data management procedures. Data collected by NPS staff, cooperators, researchers and others must be entered, quality-checked, analyzed, reported, archived, documented, cataloged, and made available to others for management decision-making, research, and education. This Data Management Plan (DMP) serves as the overarching strategy for achieving these goals. The plan supports I&M Program goals and objectives by ensuring that Program data are documented, secure, and remain accessible and useful indefinitely.

Great Lakes Network

The Great Lakes Network (GLKN or Network) is composed of nine park units in the western Great Lakes region, extending from Indiana Dunes National Lakeshore in northern Indiana, to parks in Minnesota, northern Michigan and Wisconsin (Figure 1.1). The GLKN is working to build a holistic picture of change across the Network ecosystems – specifically, to detect change in ecological components and in the relationships among those components. In 2001, the GLKN began a 5-year program of biological inventories, and in 2002, began planning a Network-wide monitoring program. Each network is required to develop a formal Monitoring Plan prior to the initiation of monitoring. A DMP is a required adjunct to each network's Monitoring Plan. This document has been prepared for submittal and review on the same schedule of the Network's Monitoring Plan (Route and Elias 2005).

Natural Resource Inventories

The natural resource inventory is composed of twelve basic datasets (Appendix A) and forms the basis for planning and development of the Network's monitoring program. The biological inventories focus on vertebrate taxa and vascular plants. These inventories (*www.nature.nps.gov/im/units/GLKN/Inventory.htm*) help meet programmatic I&M goals and provide the GLKN with baseline ecological assessments from which long-term monitoring may begin.

Data resulting from the natural resource inventories and other short-term studies will be subjected to the same management standards as the monitoring data.

Figure 1.1. Park units of the Great Lakes Network. Land cover background is from the National Land Cover Dataset by the U.S. Geological Survey (from Landsat imagery circa 1990).

1.1. Context of the Data Management Plan

A wealth of data and information management guidelines has been published or is generally available via the internet. The National Park Service's I&M Program provides various guidelines (*www1.nrintra.nps.gov/im/datamgmt* – accessible via NPS computers only) to the 32 networks and parks in general. The GLKN data management strategy draws from these guidelines and formalizes them as Network policy. More detailed data management strategies are documented in standard operating procedures specific to a given data collection effort. These standard operating procedures fall under the aegis of this plan and adhere to the guidance, strategies and policies herein. Figure 1.2 depicts this data management guideline/policy inheritance model for the GLKN.

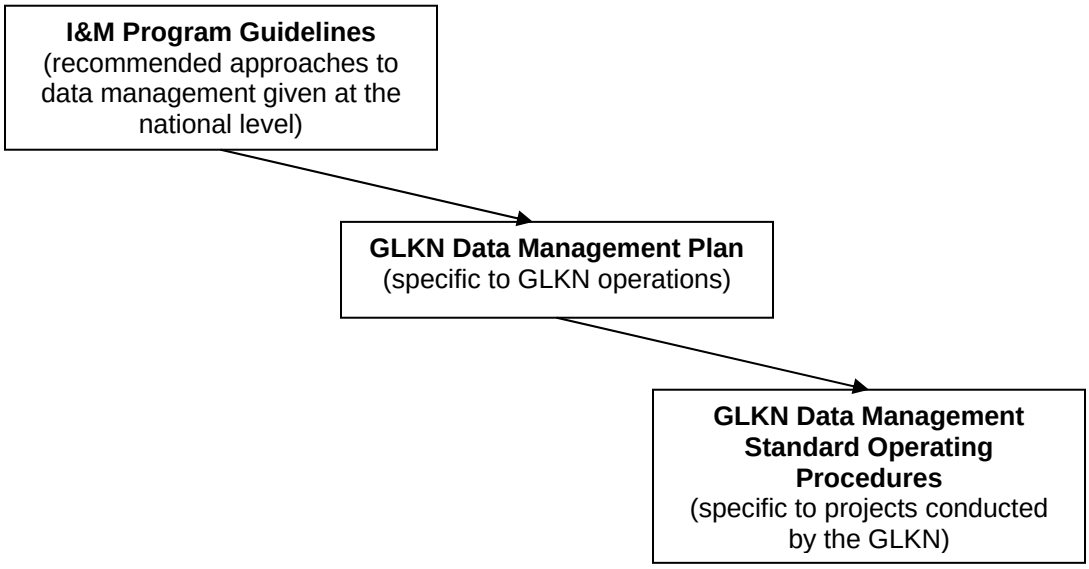


Figure 1.2. Context of National I&M Program guidelines, the GLKN Data Management Plan and detailed data management procedures for a project.

1.2. What is being managed?

Natural resource data are the vital building blocks for our evolving ecological understanding of park resources. But a set of data – whether collected the previous year or 20 years ago – must be accompanied by sufficient information on how and why it was collected to maintain its value beyond the lifetimes of those who collected it. Therefore, a data management strategy cannot simply attend to the tables, fields, and values that make up a dataset. There must also be a process for documenting the context that makes it interpretable and valuable.

The term "data" frequently encompasses other products that are generated alongside the tabular and spatial data that are the primary targets of our data management efforts. These products fall into five general categories: raw data, derived data, documentation, reports, and administrative records (Table 1.1).

Table 1.1. Categories of data products and project deliverables.

Category	Examples
Raw data	GPS rover files, raw field forms and notebooks, photographs and sound/video recordings, telemetry or remote-sensed data files, biological voucher specimens
Compiled/derived data	Relational databases, tabular data files, GIS layers, maps, species checklists, analyzed data
Documentation	Data collection protocols, data processing/analysis protocols, record of protocol changes, data dictionary, FGDC/NBII metadata, data design documentation, quality assurance report, catalog of specimens/photographs
Reports	Annual progress report, final report (technical or general audience), periodic trend analysis report, publication
Administrative records	Contracts and agreements, study plan, research permit/application, other critical administrative correspondence

Each category of project deliverables listed in Table 1.1 requires some level of management in order to place data products in the appropriate context. It is our intent to integrate the products listed in Table 1.1 such that a project's context within the overall program is maintained. This requires a holistic view of how natural resource data are generated, processed, finalized, and made available. Such an understanding allows us to tailor our data management strategies to meet our objectives in the most efficient manner possible. Chapter 5 explores more thoroughly the types of data and information managed by the GLKN and the general process by which they are generated.

1.3. Plan Scope

This Data Management Plan stipulates the broad principles for handling ecological data from their most elementary state to their final condition in reports or other formats. While this plan places heavy emphasis on data generated by, or considered important to, the GLKN monitoring program, the management principles presented may be applied to any data-gathering effort, short- or long-term, made by park resource staff. Core data management elements addressed in this plan include data collection, backup, quality control/assurance, validation, verification, documentation, archival, standard manipulation/analysis, organization, availability/access and distribution.

Document Overview:

- Chapter 1: Introduction, background information, plan overview
- Chapter 2: Outline of the data management personnel and their roles and responsibilities. Where possible, shared responsibilities are noted.
- Chapter 3: Overview of the I&M data management process. Aspects specific to the GLKN, including Vital Sign monitoring components, are addressed in context of the overall I&M Program.
- Chapters 4 through 11: More specific data management infrastructure, standards and practices related to data acquisition, quality control, and documentation.
- Appendices: Information that directly supports this document as well as supplementary material supporting data management operations in general. The latter includes standard operating procedures for transferring program data to nationally-maintained I&M databases such as NPSpecies, NatureBib and NR-GIS. It is anticipated that the appendices will experience the brunt of revisions and additions in the data management plan both as the monitoring program develops and as general Program operations persist over time.

1.4. Plan Objectives

It is important to understand the objectives of this document relative to the national I&M Program and data management goals. The goal of the NPS I&M Program is to provide scientifically and statistically sound data to support management decisions for the protection of park resources. The goal of data management is to ensure the quality, interpretability, security, longevity, and availability of ecological data and related information resulting from resource inventory and monitoring efforts. The goal of the Data Management Plan is to outline the procedures and work practices that support effective data management. The DMP will guide current and future staff of the GLKN

and ensure that sound data management practices are followed in any data gathering efforts conducted or administered by the GLKN. Where appropriate, the DMP will also guide the enhancement of legacy data to match formats and standards put forth in this plan.

It is the intention of the GLKN to establish sound data management practices such that they may be easily adopted by any resource management project. Further, where appropriate, the GLKN will promote effective data management and seek to educate personnel involved in the collection of natural resource data in the parks. By establishing robust data management and maintaining the transparency of its inner workings, the GLKN hopes to lead by example and positively affect all data management within the parks.

Data management plan objectives fall into two categories. Overall objectives describe how data management fits within the larger I&M Program, and specific objectives detail the procedures and policies necessary to implement an effective data management program.

Overall Objectives:

- Outline the long-term goals of a comprehensive data management strategy for the Great Lakes Network
- Associate data management goals with the long-term goals of the Network and Service-wide I&M Program
- Outline the procedures and work practices that support effective data management
- Guide current and future Network staff to ensure that sound data management practices are followed
- Guide the enhancement of legacy data to match formats and standards put forth in this plan
- Encourage effective data management practices as an integral part of project management so all data are made available and usable for park management decisions now and into the future
- Optimize and promote interagency sharing and development of data, software applications and analyses

Specific Objectives:

- Establish roles and responsibilities of program staff for managing data
- Identify necessary elements for a functional data management program and describe any anticipated changes to those elements
- Establish an organization schema for Network data and information so that they are retrievable by staff, cooperators, and the public
- Establish basic quality control standards
- Establish standards for data, data distribution, and data archiving to ensure the long-term integrity of data, associated metadata, and any supporting information

1.5. Plan Revisions

The plan will be formally reviewed and revised at least every 5 years. Appendix B outlines the standard revision schedule for this document. Informal review and revision will be ongoing to meet the changing needs of the program. The latest version of the DMP will be available on the GLKN website (*www.nature.nps.gov/im/units/GLKN/DataMgt.htm*) and will include a revision log as an appendix.

Credits

This chapter was adapted from concepts and material developed by Doug Wilder (Central Alaska Network). Sara Stevens (Northeast Coastal and Barrier Network) also made significant contributions to the original template version of this chapter.

2. DATA MANAGEMENT ROLES AND RESPONSIBILITIES

Data management is about people and organizations as much as it is about information technology and database theory and application. Nearly every person in an organization manages data and information at some level. Good data stewardship cannot be accomplished by data managers alone — it is truly a collaborative endeavor that involves many people with a broad range of tasks and responsibilities. The Great Lakes I&M Network contributes to the public service mission of the National Park Service and the Inventory and Monitoring Program by providing status and trends information about park ecosystems. This requires every individual involved in Network business to understand and perform data stewardship responsibilities in the production, analysis, management, and/or end use of the data (Table 2.1).

As coordinator of these activities, the Network data manager must understand Program and project requirements, create and maintain data management infrastructure and standards, and educate, communicate, and work with all responsible individuals. This chapter discusses data management roles and responsibilities that generally apply to all Network activities. Individuals who carry out monitoring protocols and inventory study plans are responsible for reading and understanding guidance at two levels — the overarching level of this Data Management Plan and the specific, task-oriented level of the respective protocol or study plan. This chapter clarifies roles and outlines the principal and ancillary data management responsibilities of Network staff, cooperators, and other project participants.

Table 2.1. Data stewardship categories for all Network personnel and cooperators.

Stewardship Category	Related Activities	Position*
Production	Collecting data or information from any original or derived source such as recording locations, images, measurements, and observations in the field; digitizing source maps; keying in data from a hardcopy source; converting existing data sources; image processing; and preparing and delivering informative products, such as summary tables, maps, charts, and reports.	Project Crew Leader Project Crew Member Project Data/GIS Specialist or Technician
Analysis	Using data to predict, qualify, and quantify ecosystem elements, structure, and function as part of the effort to understand these components, address monitoring objectives, and inform park on ecosystem management.	Network Ecologist Park Resource Specialist Network Statistician
Management	Preparing and executing policies, procedures, and activities that keep data and information resources organized, available, useful, compliant, and safe.	Network Data Manager Project Manager Regional GIS Manager Regional IT Specialist Project Database Manager Service-wide I&M Data Manager
End Use	Obtaining and applying available information to develop knowledge that contributes to understanding and managing park resources.	Network Coordinator Park Manager Superintendent Other

**NOTE – Each position is listed in only one category according to primary responsibilities; however, most positions contribute to activities in each category.*

2.1. Data Stewardship Roles and Responsibilities

Definitions

- A role is a function or position (e.g., project manager)
- A responsibility is a duty or obligation (e.g., review data records)

An increasing demand for detailed, high quality data and information about natural resources and ecosystem functions requires a group of people working together to steward data and information assets. Knowledgeable individuals from scientific, administrative, and technological disciplines must work in concert to ensure that data are collected using appropriate methods, and that resulting datasets, reports, maps, models, and other derived products are well-managed. Datasets and the presentations of these data must be credible, representative, and available for current and future needs. Stewardship responsibilities apply to all personnel who handle, view, or manage data (Table 2.2). Vital Sign monitoring protocols will describe more detailed project management and data stewardship roles and responsibilities.

Table 2.2. Programmatic roles and associated data stewardship responsibilities.

Role	Data Stewardship Responsibilities
Project Crew Member	Record and verify measurements and observations based on project objectives and protocols. Document methods and procedures.
Project Crew Leader	Supervise crew members to ensure adherence to data collection and data processing protocols, including data verification and documentation.
Database Manager (Project)	Apply particular knowledge and abilities related to database software and associated application(s).
Data/GIS Specialist or Technician (Project or Network)	Support the Network Data Manager and others in implementing data management procedures. Duties may include GIS activities, cataloging, data conversion and integration, documentation, and reporting.
Statistician/Biometrician/ Quantitative Ecologist	Develop project sampling design. Analyze data, develop models, and documentation procedures. Review methods for statistical soundness.
Project Manager	Direct project operations. Communicate data management requirements and protocols to project staff, Network Data Manager, and resource specialist(s). Responsible for final submission and review of all products and deliverables.
Park Research Coordinator	Facilitate data acquisition by external researchers. Communicate NPS requirements to permit holders.
Park Resource Specialist	Understand project objectives, data, and management relevance. Make decisions about validity, sensitivity, and availability of data.
Curator (Park or Region)	Manage collection, documentation, and preservation of specimens.
Network Data Manager	Oversee development, implementation, and maintenance of data infrastructure and standards. Facilitate and integrate data and metadata. Oversee long-term data storage and maintenance. Design and develop databases and applications.
Network Ecologist	Ensure useful data are collected and managed by integrating natural resource science into Network activities and products, including specifying objectives, sample design, data analysis, synthesis, and reporting.
Network Coordinator	Ensure programmatic data and information management requirements are met as part of overall Network business.
GIS Manager (Network or Region)	Provide support for long-term storage of GIS data. Update and maintain GIS software and tools. Provide technical assistance.
Information Technology Specialist (Network or Region)	Maintain LAN, establish and maintain system security, update software and hardware, and implement secure file server backup system.
I&M Data Manager (National)	Provide Service-wide database design, support, and services, including processing to convert, store, and archive data in Service-wide databases.
End Users (managers, scientists, interpreters, public)	Provide feedback on scientific information, presentation needs, and interpretation. Use information for management decisions.

2.2. Project Stewardship

Project managers will direct data collection for Vital Signs monitoring and/or inventory efforts. They will provide project oversight, direct on-the-ground data collections, and provide a cohesive link between data collection, synthesis, interpretation, and reporting. Project managers must act as stewards for all project data and must work with project and GLKN personnel to ensure that project data are handled properly.

2.2.1. The Importance of Documentation

Among all of the data management responsibilities shared by project and Network personnel, the careful documentation of datasets, data source(s), and data collection methodology is paramount. This careful and thorough documentation establishes the basis for the appropriate use of the data in resulting analysis and reporting, both in the short-term and long-term. Network monitoring protocols contain key elements of data documentation. Network data records collected according to these protocols will include the name, date, and version of the associated protocol. (Chapter 7 presents important guidance and reference for documentation and metadata.)

2.3. Data Stewardship Team

Network coordinators, project managers, data managers, GIS specialists, and other data specialists comprise the central data management team for inventory and monitoring projects. Each is responsible for certain aspects of project data and all share responsibility for some overlapping tasks (Figure 2.1). Because of the collaborative nature of project data management, good communication among these personnel is essential in order for program goals to be met.

Communication is promoted by providing:

- Documentation
- Shared working files
- Centralized data management guidance and information discovery on the GLKN website
- User needs assessments and surveys
- Newsletters and periodic e-mail updates
- Work groups discussions and meetings
- Presentations
- Training

2.3.1. Network Coordinator

The network coordinator supervises the project managers and has the ultimate responsibility for data entry, validation, verification, summarization/analysis, and reporting. The Network's Board of Directors and Technical Committee give final approval for products to be integrated, distributed to the public or parks, or protected, in the case of sensitive information.

2.3.2 Project Manager

Project managers oversee and supervise all phases of an inventory and monitoring project from initiation to product delivery. Assuming 'ownership' of the data, the project manager's active involvement in data management determines the quality and usefulness of the project data and contributes to the overall success and longevity of the I&M Program. Project managers are responsible for designating an alternate leader who is capable of maintaining project operations in his or her absence, in order to ensure project continuity and data integrity.

Specifically, a project manager is responsible for:

- project documentation that describes the 'who,' 'what,' 'where,' 'when,' 'why,' and 'how' of a project
- documentation and implementation of standard procedures for collecting and handling field data
- quality assurance and quality control measures, which include the supervision and certification of all field operations, staff training, equipment calibration, species identification, data collection, data entry, verification, and validation
- maintenance of concise documentation explaining all deviations from standard procedures
- detailed documentation of post field-collection methodology
- maintaining and archiving of original and hard copy data forms
- scheduling of regular project milestones such as data collection periods, data-processing target dates, and reporting deadlines
- producing regular summary reports, periodic analysis of data, resulting reports, and making reports available to the public
- identifying sensitive information that requires special consideration prior to distribution
- acting as the main point of contact concerning data content/quality

The project manager may also work closely with the data manager and/or other data personnel (such as a biometrician) to:

- develop quality assurance and quality control procedures specific to project operations
- identify training needs for staff related to data management and quality control procedures
- coordinate changes to the field data forms and the user interface for the project database (coordinate data entry procedures—data design maintenance)
- document and maintain master data (including metadata generation and maintenance)
- manage the archival process to ensure regular archival of project documentation, original field data, databases, reports and summaries, and other products from the project
- define the process of how project data will be transformed from raw data into meaningful information
- create data summary procedures to automate and standardize this transformation process

- identify and prioritize legacy data for conversion and convert priority datasets to a modern format
- increase the interpretability and accessibility of existing natural resource information
- catalog project data and reports in nationally maintained I&M Program databases such as NPSpecies, NatureBib, and NR-GIS Data Store (see Chapter 10).

2.3.3. Network Data Manager

The data manager oversees the development, implementation, and maintenance of data infrastructure and standards for the Network. The data manager is responsible for ensuring that project data are compatible with program standards and for the long-term integrity and availability of project data. Working with project managers, GIS specialists, and other Program staff to design databases, applications, and products, the data manager facilitates dissemination of project datasets and information products. The data manager ensures that data are archived, documented, and compatible with other program data, and sees that data are discoverable and available using appropriate mechanisms. Both the data manager and the project manager ensure that the information conveyed via these mechanisms is up-to-date and accurate.

General data management duties for the data manager are:

- overall coordination of data management activities in the Network
- developing and maintaining systems to house and disseminate data and information for all programs
- improving the acquisition, accessibility, and transparency of digital data and ensuring that the data and information system is populated and kept up-to-date with all relevant Network output
- developing and maintaining logs recording both the changes and enhancements to the data and information-handling system and processes
- acting as point of contact for access to Network output
- ensuring data security (archiving operations, etc.)

The data manager will also work closely with the project managers to:

- develop and maintain the infrastructure for metadata creation, project documentation, and project data management
- create and maintain project databases in accordance with best practices and current program standards
- provide training in the theory and practice of data management tailored to the needs of project personnel
- establish and implement procedures to protect sensitive data according to project needs
- collaborate with GIS specialists to integrate tabular data with geospatial data in a GIS system in a manner that meets project objectives
- define the scope of the project data and create a data structure that meets project needs
- become familiar with how the data are collected, handled, and used

- review quality control and quality assurance aspects of project protocols and standard procedure documentation
- identify elements that can be built into the database structure to facilitate quality control, such as required fields, range limits, pick-lists, and conditional validation rules
- create a user interface that streamlines the process of data entry, review, validation, and summarization that is consistent with the capabilities of project staff
- develop automated database procedures to improve the efficiency of the data summarization and reporting process
- make sure that project documentation is complete, complies with metadata requirements, and enhances the interpretability and longevity of the project data
- ensure regular archival of project materials
- inform project staff of changes and advances in data management practices

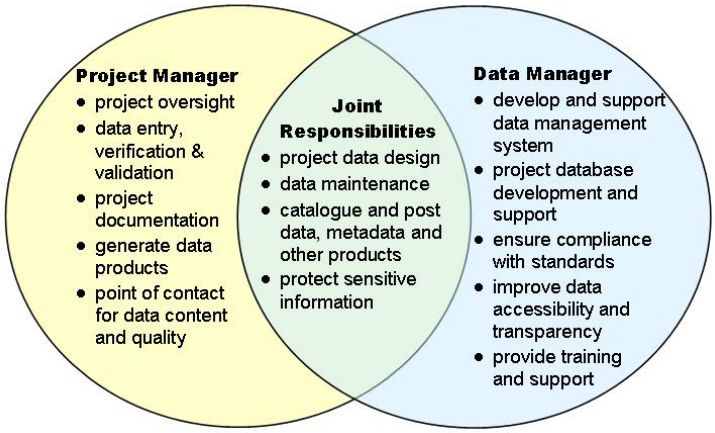


Figure 2.1. Core data management responsibilities for project and data managers.

2.3.4. Regional and Network GIS Specialists

The Regional GIS specialist and the Network GIS specialist play a crucial supporting role for data management of GLKN projects. GIS support includes project planning to determine the GIS data and analysis needs for a project. GIS specialists advise project managers on methodology related to field collection of spatial data (i.e., use of GPS and other data collection techniques). They will also coordinate importing of spatial data into the GIS, work with project managers to analyze spatial data, and provide the results in map or tabular form. The GIS specialists will work with project managers to properly document data in compliance with spatial metadata standards and will be responsible for stewardship of GIS data and products.

GIS specialists will also work directly with data managers to design databases and applications, to create relationships between GIS and non-spatial data, and to create appropriate database and GIS applications to facilitate the analysis of both spatial and non-spatial data. Maintaining standards for geographic data, GIS specialists are responsible for sharing and disseminating GIS data throughout the Network.

2.4. Data Management Coordination

The Natural Resource Challenge states that collaboration among the National Park Service, other public agencies, universities, and non-governmental organizations is necessary to effectively acquire, apply, and promulgate the scientific knowledge gained in national parks. The I&M Program encourages coordination among participants at all levels to help ensure that data collected in parks are properly developed, maintained, and made available for management decision-making, research, and education.

The Network data manager works with I&M Program data management staff and regional resource information management personnel to maintain a high level of involvement in data management policy and in maintenance of Service-wide and regional databases. The data manager also works locally with Network personnel, park staff, and cooperators to promote and develop proficiency with Service-wide databases and workable standards and procedures that result in the compatibility and availability of datasets.

At the Network level, key contacts include GIS specialists and the project managers for each monitoring or inventory project. Involvement and input from park scientists, GIS specialists, and resource management staff is also essential. We rely on everyone in the Network for the successful development of planning materials, inventory study plans, and monitoring protocols. Consistent and productive communication among these individuals leads to common understanding and better synchronization of data management activities. This communication may take the form of personal visits, phone calls, email, joint meetings and training sessions, and participation in meetings of the GLKN's Technical Committee.

Data managers throughout the NPS regularly coordinate with each other and national program staff via annual meetings, conference calls, workgroups, a listserv, web sites, and informal communication. Data managers have already demonstrated effective cooperation in the ways that we share ideas and technology, and collaborate to develop data management standards and documentation. This communication and cooperation promotes compatibility of protocols and datasets among networks and organizations.

Credits

This chapter was adapted from concepts and material developed by Rob Daley (Greater Yellowstone Network) and Doug Wilder (Central Alaska Network).

3. DATA MANAGEMENT OVERVIEW AND WORKFLOW

The infrastructure, procedures, and objectives of data management must conform to project operations as well as overall I&M Program operations in order to be accepted and effective.

The project process flow presented in this chapter is idealized. It is recognized that this flow and related data management strategies and infrastructure may not apply in all situations; however, planning for good data management requires an understanding of how work is or will be conducted in the Network. Where that understanding is not possible because project operations are still in development, this plan takes the liberty of defining the process work flow.

An understanding of the common data-related products and procedures is also necessary for successful implementation of this plan. Base operating procedures and the array of anticipated products for the GLKN are still under development at this time (plan submitted 12/15/2005). To the extent necessary to meet the goals of this plan, those procedures and products are surmised in the context of data management.

3.1. General Project Work

To assure effectiveness, the data management process must be pervasive in the execution of a project. At an operational level, the process used to implement any data-gathering project should be clearly defined and governed by protocols that include relevant data management elements. Data management goals and directives need to integrate with the workflow design and procedures developed by the program managers of the various monitoring protocols. Data management processes must be relatively transparent and seamless from the perspective of the ecologists and researchers who need to access and use the data.

A generalized project work flow model provides the framework for data management. While the work flow presented in this plan may not apply to all situations, it does address both the long- and short-term data collection efforts anticipated by the GLKN monitoring program.

A project is divided here into the following stages:

- Project initiation
- Planning and approval
- Design and testing
- Implementation
- Product delivery
- Product integration
- Close-out and evaluation

Project stages and related activities are depicted in Figure 3.1. Appendix D offers details of anticipated deliverables, deliverable repositories, and references to data management guidance documents for each stage.

As part of its monitoring program, the GLKN will use a project tracking database, a slightly modified version of a database developed for the I&M Programs in Alaska. To the extent feasible, information in this project tracking database will be mapped to nationally maintained systems as appropriate. The GLKN project tracking database will support coordination of the monitoring program and annual reporting. This project-tracking database will be modified as necessary as the monitoring program develops.

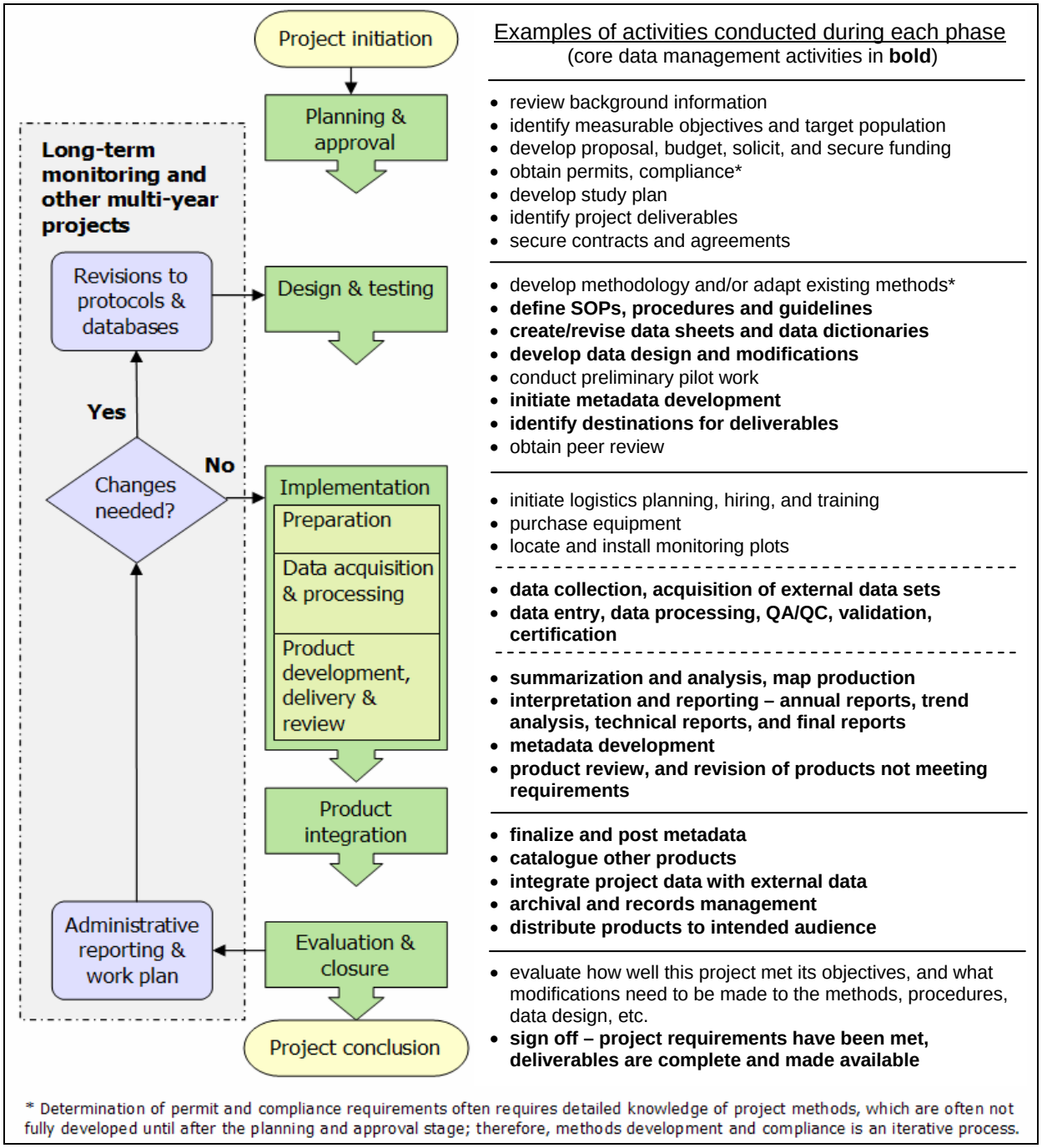


Figure 3.1. Generalized project work flow model for short- and long-term data collection efforts. Bold face activities indicate core data management elements.

3.2. Core Data Management Elements

Essential data and information management includes planning for data collection, actual collection, data entry, validation (quality control, etc.), and submittal to designated archives and audiences. Project activities involving these elements are highlighted in Figure 3.1.

As indicated in Figure 3.1, metadata should develop as the project progresses rather than after data are collected, entered, validated, verified, and submitted. Similarly, quality control of data should begin with the data design. Field sheets and databases should be constructed to minimize, and where possible, restrict the variety of values that may be recorded for a given parameter. Pick lists, whether built into a data logger or included on a field sheet, should be employed wherever possible.

Certain key information is not only common to datasets, but to the organization as a whole – lists of contacts, projects, parks, and species are often complex and dynamic. It is a good strategy to centralize such information so that users have access to the most updated versions in a single known place. Centralization also avoids redundancy and versioning issues among copies. Centralized information is maintained in database tables that can be linked or referred to from several distinct project databases. Network applications – for project management, administrative reporting, or budget management – can also link to the same tables so that all users in the Network have instantaneous access to edits made by other users.

The GLKN will also utilize common tables for a variety of parameters collected by different Vital Sign projects. The national I&M Program's Natural Resource Table Standards (NRTS) is an effort to standardize commonly collected data (e.g., percent of cloud cover, precipitation intensity) and will be used by the GLKN.

The National Park Service document, *Recommended Database Strategies* (Southwould 2002), also contains common data practices that the GLKN will adopt for Vital Sign monitoring. The document is available at *science.nature.nps.gov/im/datamgmt/links.htm*.

Credits

This chapter was adapted from concepts and material developed by John Boetsch (North Coast and Cascade Network), with input from many others.

4. DATA MANAGEMENT INFRASTRUCTURE

Data management infrastructure refers to the complex, underlying system that supports data management. This system includes hardware (computers and servers) that are functionally or directly linked through computer networking services. While networked hardware provides the foundation on which the Network information system is built, systems architecture (the applications, database systems, repositories, and software tools used in our data management enterprise) is an equally important part of the infrastructure. Therefore, the design of mission-critical databases is part of the logical infrastructure of the Network.

The GLKN monitoring program relies heavily on park, regional, and national information technology (IT) personnel and resources to maintain the computer resource infrastructure. Ongoing maintenance of computer resources includes, but is not limited to, hardware replacement, software installation and support, security updates, virus-protection, telecommunications networking, and backup processes. Communication with park, regional and Network IT specialists is essential to ensure service continuity for the Network's systems architecture. Rather than focusing on a detailed description of the current computer resources, this chapter will focus on the systems architecture that is central to data management.

At this time, the GLKN data management system is in a developmental stage. The system design presented here represents current plans as monitoring protocols are developed and implemented.

This chapter also specifies standards for GLKN data management operations. In general, the GLKN will conform to National Park Service standards and policies in all aspects of Program data management operations. Further, the GLKN will conform to national I&M Program standards and mandates in the interest of program integration and information sharing.

4.1. NPS Infrastructure – Components

The Great Lakes Network has invested significantly in a reliable, secure computer network system. The Network also shares an IT Specialist position with Apostle Islands National Lakeshore. The IT Specialist contributes substantially to the smooth and consistent operation of the Network's computer systems.

The Great Lakes Network digital infrastructure consists of two in-house servers connected to workstations through the Regional Office using Microsoft's Active Directory architecture. The Network currently uses a cable connection to access the NPS WAN and Internet, but may switch to a T1 line if increased bandwidth becomes necessary.

Databases and information are stored on the servers, which are backed up to tape daily. Each week, a backup tape is transferred to an offsite location. Workstation backups are also run to tape each week.

The Network maintains a consistent LAN/WAN connection in order to access data and information stored on remote computers. Distributed data are accessed through NPS

servers, which host applications such as NatureBib, NPSpecies, and the NR-GIS Metadata and Data Server, and through sites maintained by partner agencies (e.g., the National Weather Service, USGS or EPA).

There are distinct advantages to this distributed data architecture. Data are maintained and archived by the stewards of those datasets, the data provided is consistent in quality and format, and local hard drive space is not filled with outdated copies of the information.

The Network is developing an ArcIMS site in cooperation with the national I&M Program which park staff, cooperators, and researchers will use to access Network data (see Chapter 9). This application will allow for data exploration, tabular and spatial queries, and download capabilities. We also plan to upload monitoring data directly to this site from the field when possible. By providing access to the most current data, the Network will eliminate the need for manual extraction and dissemination and will minimize versioning issues.

The GLKN staff stationed at the office in Ashland, Wisconsin utilizes the Network LAN for data exchange and common storage. The configuration and components in the LAN will change as equipment is replaced and as new technology becomes available. It is also probable that Network staff will access LAN resources over the Service-wide WAN in the future. At the time this document was developed, the GLKN LAN was comprised of the following components:

- Workstation/laptop
- PowerEdge 6600 400 Gb Raid 5
- PowerEdge 2850 670 Gb
- PowerVault 122T Backup system
- Backups (server daily, laptops weekly)

The following list outlines the data components that encompass the distributed data within the NPS I&M Program for the GLKN. The park LAN components are not a resource that is maintained specifically for use by the Network, but they will be used by Network staff and cooperators in the course of research activities.

4.1.1. National servers

The national server components are:

- Master applications – integrated client-server versions of NatureBib, NPSpecies, NR-GIS Metadata Database
- Centralized repositories – NR-GIS Data Store, Protocol Clearinghouse
- Public access sites – portals to NatureBib, NPSpecies, NPSFocus, and websites for monitoring networks

4.1.2. Network data server

The network server components are:

- Master project databases – compiled datasets for monitoring projects and other multi-year efforts that have been certified for data quality
- Common lookup tables – e.g., parks, projects, personnel, species

- Project tracking application – used to track project status, contact information, product due dates
- Network digital library – Network repository for finished versions of project deliverables for Network projects (e.g., reports, methods documentation, data files, metadata, etc.)
- Local applications – desktop versions of national applications such as NPSpecies and Dataset Catalog
- Working files – working databases, draft geospatial themes, drafts of reports, administrative records
- GIS files – base spatial data, imagery, and project-specific themes

4.1.3. Network computers

The network computer components are:

- Draft project databases – databases in development
- Working files – Draft reports, administrative records
- Draft GIS files – Projects in development
- Local applications – desktop versions of national applications such as NPSpecies and Dataset Catalog

4.1.4. Park LAN

The park LAN components are:

- Local applications – desktop versions of national applications such as NPSpecies and Dataset Catalog
- Working files – working databases, draft geospatial themes, drafts of reports, administrative records
- Park digital library – base spatial data, imagery, and finished versions of park project deliverables
- Park GIS files – base spatial data, imagery, and project-specific themes

The components of the information system architecture will change as the I&M Program matures. Current national-level I&M data management infrastructure and strategy is presented in Appendix C. Additions to existing regional-, Network- and park-level infrastructure will be made in order to meet data management objectives.

Two regional programs contribute to the Network's data management infrastructure. The Midwest Regional Office (MWR) provides software updates, some hardware upgrades, and technical support related to security and networking issues. The Midwest Regional GIS Office distributes software updates for ESRI's ArcGIS products and helps develop and complete GIS projects for the Network and MWR parks.

The GLKN infrastructure is composed of elements from the national I&M Program, the MWR and GLKN parks. Figure 4.1 illustrates the connectivity between national, Network, and park information systems. The Network utilizes the national infrastructure outlined in Appendix C for its public website presence and for file-sharing via the NPS FTP site.

In addition to maintaining the in-house client/server architecture, the GLKN is in the process of implementing an enterprise GIS system architecture using ArcIMS/SDE with a server-based SQL relational database management system (RDBMS).

There are many advantages to implementing this system at the Network level. The complexity of numerous databases for the various Vital Signs monitoring programs for nine parks as well as the large raster datasets of imagery warrants an enterprise architecture for increased efficiency in storage and retrieval of data, database security, maintenance, and versioning. Our plan is to design enterprise geodatabases and take advantage of the efficiencies of employing ArcSDE with an RDBMS for data integrity, versioning, disconnected editing, and more rapid data transactions.

Chapters 9 and 10 of this plan present more details on the distribution, maintenance, and storage of GLKN data. These chapters may be summarized as:

- The GLKN will manage a primary repository located in the Network office for data and information generated by the Network.
- A portion of the GLKN SQL data tables required for the efficient operation of the Network's IMS gateway may be mirrored on WASO servers, and access restricted to authorized users.
- The primary GLKN repository will be backed up, and media stored offsite locally, with periodic archives of the information to a secure facility contracted at the WASO level.
- Required portions of the GLKN data will be stored in national I&M Program databases such as NPSpecies, NatureBib, and NR-GIS Data Store.
- Certain GLKN data sets will be primarily maintained by outside organizations; however, metadata for these data sets will be maintained in the primary GLKN repository. An example of this category is the climate data which will be handled by National Oceanic and Atmospheric Administration (NOAA) or a designee, such as the Midwest Regional Climate Center, under formal agreement with the NPS and/or the GLKN.

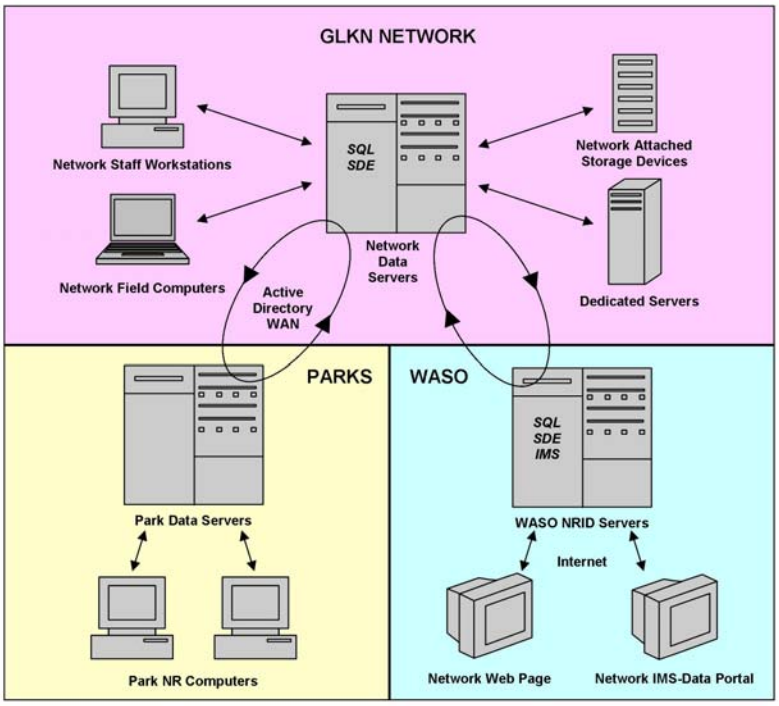


Figure 4.1. Information technology connectivity between Network, Park and National operations.

4.2. Data Management Hardware Standards

The GLKN information management hardware will conform to NPS standards and policy. Where possible and appropriate, the GLKN will take advantage of new developments in computer technology to enhance data collection, quality control, storage, and accessibility.

4.3. Basic Software Standards

The software used by the GLKN will conform to NPS standards and policy. The subsections below provide further detail regarding current standard practices within the GLKN. As products and technologies change, the GLKN will evaluate potential changes in software platforms, with active consideration given to maintaining access to existing file formats.

4.3.1. Word Processing (Microsoft Word and Adobe Acrobat)

All reports and other textual documents will be finalized in editable electronic format in the latest available version of Microsoft Word or Adobe Acrobat, unless otherwise specified in the project study plan. Distribution copies will be converted to the latest version of Adobe Acrobat.

4.3.2. Databases (Microsoft Access and SQL Server)

The application used for database development will depend on the scale of the project. Desktop versions of GLKN databases will be in the latest Microsoft Access format unless otherwise specified in the project study plan. Desktop databases produced under contract

will also use Microsoft Access unless otherwise specified in advance. For complex, interactive databases, Microsoft's SQL Server will be required. The GLKN is developing a server-based relational database which will serve as the main repository for Network output and facilitate data and information dissemination (see Chapter 9).

4.3.3. GIS Products (ArcGIS)

All GIS products must be compatible with ESRI's ArcGIS software. GIS products should meet NPS specifications (*www.nps.gov/gis/data_standards/DataStandards.html#Stewardship*) and, when appropriate, will be archived in uncompressed TIFF and PDF files. All GIS products will be accompanied by FGDC-compliant metadata.

4.4. Data Management System

Section 3.1 presented an overall project process flow as a framework for data and information management. Core date management practices were presented in the context of the project work flow in Figure 3.1. The major data management elements in the process include:

- Collect or acquire data (field work, satellite download, etc.)
- Archive raw data (physical and possibly digital material)
- Enter or import data
- Verify and validate data (quality assurance and quality control)
- Produce documentation (metadata, etc.)
- Archive validated data & documentation (digital)
- Integrate project data with external data (NR-GIS, NatureBib, etc.)
- Summarize and analyze data, and produce reports
- Archive data products
- Catalogue data products

The elements listed above comprise the data life cycle for most GLKN projects. Figure 3.1 presents the relative timing of each element during a project. Procedural details will be documented in the standard operating procedures for each Vital Sign. Common data management procedures are included in this plan.

4.5. Data Access Applications – Getting at the data and information

It is impossible to wholly predict the demands that will be made on a data interface for the complex body of data and information the GLKN intends to manage. However, the following basic capabilities are deemed important:

- Robust data browsing:
  - Ready access to summary results from any one Vital Sign measure
  - Ability to compare results between Vital Signs
  - Ready access to monitoring results by time and location
- Simple data access (download):
  - Easy data subsetting (data searches that allow data to be selected by date, location, subject, etc.)

  - Standard file format choices
  - Dissemination tracking
  - Accompanying metadata
- Easy data upload for project managers and authorized personnel
- Comprehensive data discovery so that our data and information may be easily found by outside stake holders

To facilitate timely and easy access to Vital Signs monitoring data sets, the GLKN is continuing the development of an ArcIMS/SDE data gateway in order to accomplish the goal of delivering information to NPS staff and cooperators. An enterprise license agreement with ESRI allows the Network to access these software products without additional overhead expense. Although considerable resources are needed to customize the application, we believe this will best meet the Network's needs.

4.6. Database Design

Although there are currently no Service-wide standards on database design, the I&M Program recognized early on that database standards and detailed documentation were needed. Database standards promote compatibility among data sets and allow data to be aggregated and summarized. Well thought-out standards help to encourage sound database design and facilitate interpretability of data sets.

The I&M Program has developed a series of recommendations in database design which include:

- Database specifications for I&M Studies
- Recommended database strategies, including the I&M Database Template
- Recommended naming standards
- Natural Resource Database Template data standards

Each database must ultimately meet the needs of the Network scientists. These needs may include compatibility with databases produced by other agencies and ease of use, maintenance, integration, and customization. The GLKN will model and document databases as described in this Data Management Plan, will use the above recommendations in database design as a guideline where feasible, and will standardize attributes across databases where feasible.

Communication is a vital part of designing a project database. Development of data models can promote communication and collaboration between project managers and data managers. Data modeling is an iterative, three-stage process. Each stage results in an increasingly complex, detailed model. The three models, which combine diagrams and associated descriptions are conceptual, logical, and physical data models.

4.6.1. Conceptual Data Models

Conceptual data models (CDM) are constructed to graphically portray processes related to project implementation – especially data acquisition, processing, QA/QC, and data reduction (Figure 4.2). These conceptual models, which are software-independent and free of database details, focus on capturing the information needed to accurately express the project data design. Conceptual data models are often created as the precursor to

logical data models (LDMs) and may be discarded after the LDM is complete. For less complex projects, CDMs may be kept as alternatives to LDMs.

Conceptual Data Models should contain the following:

- A short description in layman's terms of what is going to happen. Concise, yet descriptive, information should be included to establish a context for the database (e.g., environmental conditions while collecting or skill level of staff)
- A flow diagram of procedures, identifying what data is needed and when, and what information is produced and when.
- A description or mock-up illustration of how the data should be presented.

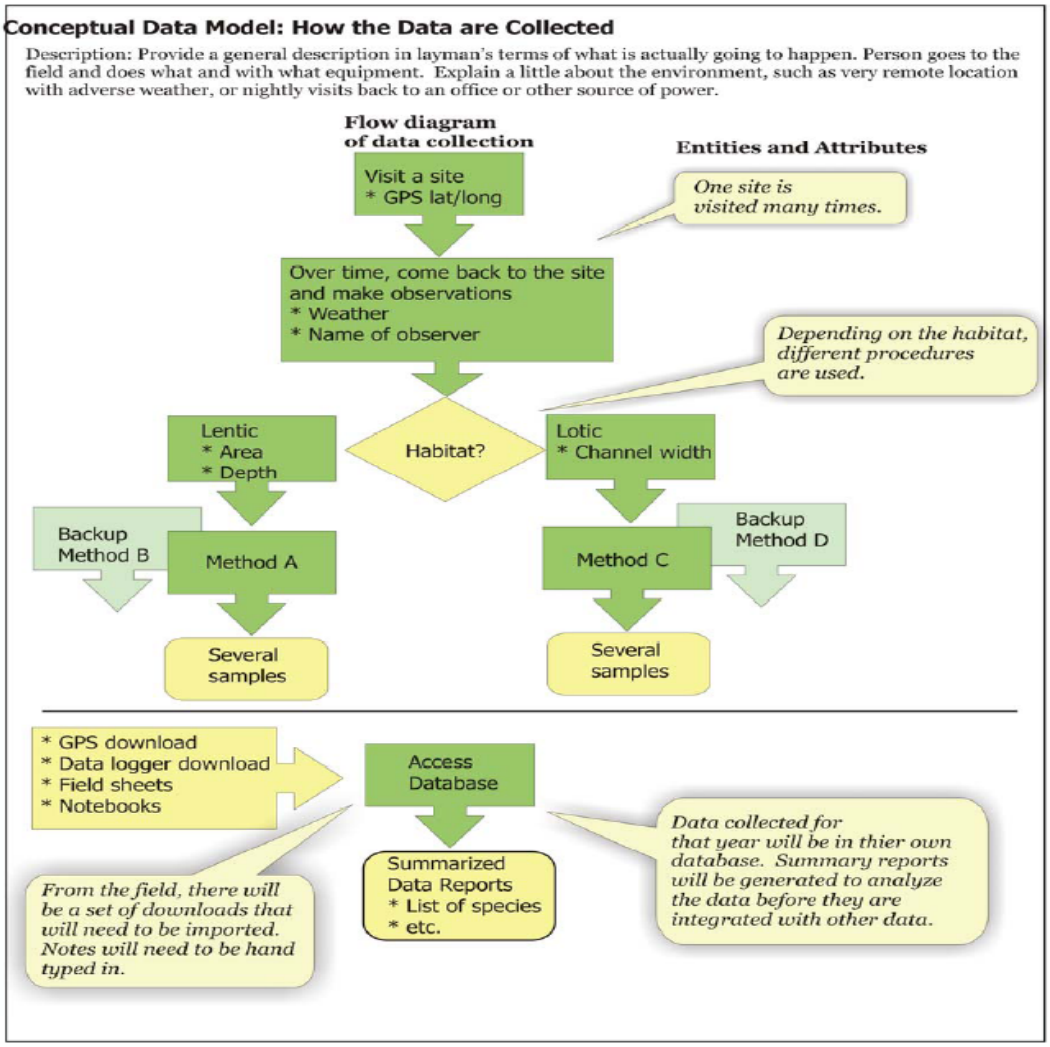


Figure 4.2. Schematic of conceptual data model.

4.6.2. Logical Data Models

A logical data model (LDM) is an abstract representation of a set of data entities. Key attributes and the relationships between data entities are identified (Figure 4.3). The logical data model is not intended to be a full representation of the physical database, but rather should facilitate analysis of the data design.

Logical data models are made up of five main components:

1. *Data entities* are distinct features, events, observations, and objects that are the building blocks of a data set.
   - sample sites
   - sampling events
   - sampling units (transects, plots, etc.)
   - watersheds
   - species
   - habitat types
   - species observations
   - tissue specimens
2. *Entity attributes* are properties and rules associated with data entities. For example,
   - Sample sites have dimensions and a geographic position
   - Vegetation transects are 100 meters long
   - Temperature readings are recorded in Celsius to the nearest tenth degree
   - Elevation is recorded to the nearest foot, and cannot exceed 9,000 feet
   - Species abundance is recorded in terms of projected horizontal cover of all aboveground parts, as estimated by trained observers. Percent cover is estimated to the nearest whole number, ranging from 0 to 100%.
   - The degree to which vegetation obstructs the field of view around animal groups is classified in three categories: high (>75%), medium (25-75%) and low (<25%)
3. *Logical relationships* explain how data entities are logically related. For example,
   - Each site will be visited numerous times
   - Each sampling event might have zero or numerous species observations
   - Each species can only be observed once per sampling event
   - Every sample must use one of three known sample methods
   - Every time a water sample is collected, temperature, pH, and dissolved oxygen must also be measured
4. *Structural hierarchies* identify the structure and order of relationships between data entities, which can be determined once the logical relationships are known. A simple example of a structural hierarchy is:
   - Site locations
     - sampling event
       i. species observations
       ii. water samples, temperature, pH, dissolved oxygen
5. *Views* identify how the data will be viewed or what operations will be performed on the data. For example,
   - Summary list of bird species per park
   - Monthly average air temperature, wind direction, and precipitation

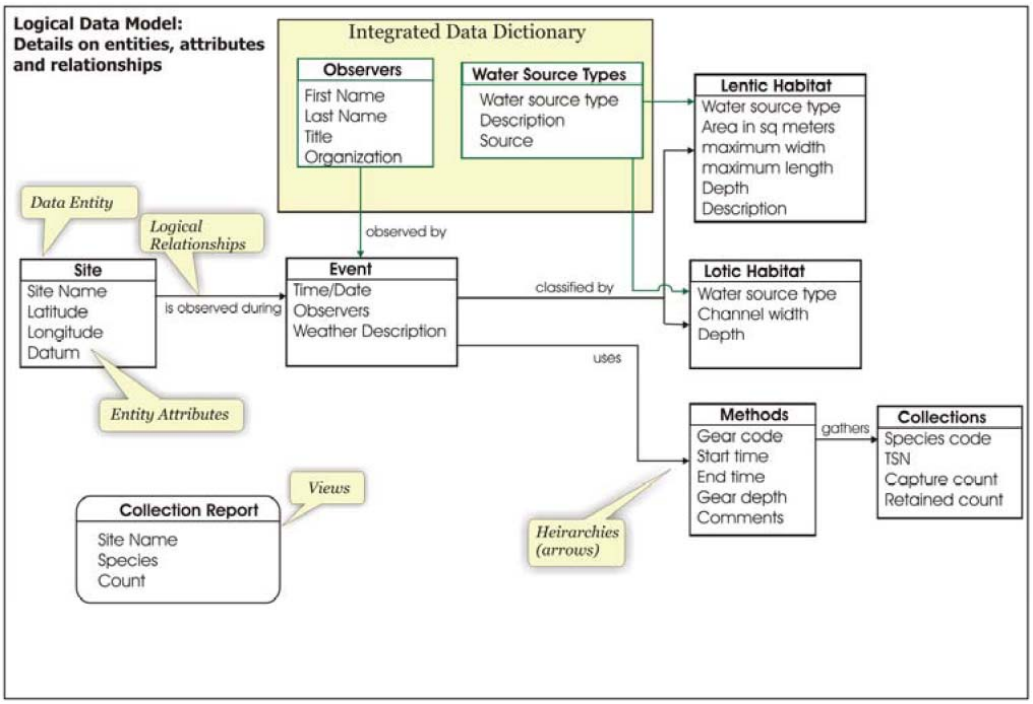


Figure 4.3. Schematic of logical data model.

4.6.3. Physical Data Models

The physical data model (PDM) is used to design the actual database and depicts data tables, fields and definitions, and relationships between tables (Figure 4.4). Although the logical and physical data models are similar, the logical data model only provides enough detail to communicate what information will be stored in the database. The physical data model provides very specific details and definitions (e.g., primary keys and field types).

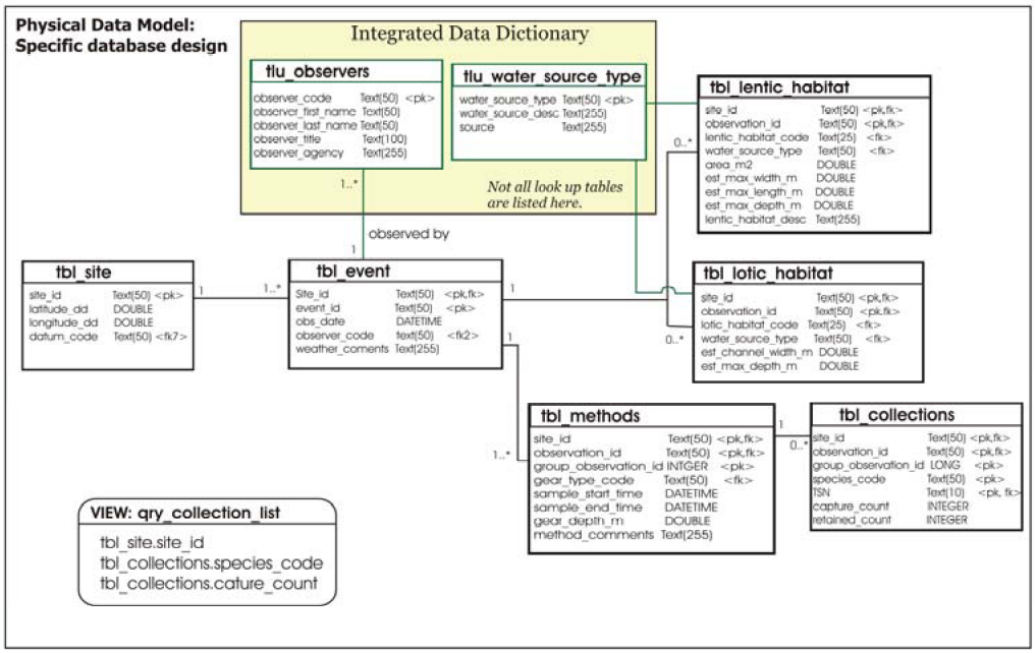


Figure 4.4. Schematic of physical data model.

4.6.4. Developing the Models

Four elements are required to develop the data models:

1. People – People will need to be involved in the model development. Discussions may begin with a few individuals and expand to a larger group as models are reviewed. Alternatively, a large group might generate the concepts and a smaller group might work out the details. The people most typically involved in data modeling are the network coordinator, scientists or project managers, partners, field crews, biometricians, and data managers.
2. Protocols – The protocols for the Vital Sign will provide the greatest substance to the models. The protocols provide the goals, objectives, methods, standards, analysis, and reporting.
3. Reference materials – Reference materials such as field forms, drawings, mock-up reports, and references to classifications to be used will play a significant role in the data models.
4. Frequent interactions -- Frequent discussions between the project leader, the data manager, and others will be needed to develop successful data models. Detailed review of the protocols and reference materials will articulate the entities, relationships, and flow of information.

Data models are not flexible by themselves. Data modeling needs to be iterative and interactive, and processes and techniques that support change should be adopted.

4.6.5. Fundamental Database Structure

The Natural Resource Database Template (NRDT) provides an example of the fundamentals for database structure to be used by the I&M Program. These fundamentals are as follows:

1. Mandatory (core) tables and fields – These tables and fields are used to manage the information describing the "who, what, where, and when" of project data. These tables contain critical data fields that are standardized with regard to data types, field names, and domain ranges. For example: tbl_Sites, tbl_Locations, tbl_Events.
2. Mandatory if Applicable (master look-up elements) tables and fields – Described in more detail below (Integrated Data Dictionary), these tables contain commonly used attributes, such as park codes, and are centrally located to minimize redundant information in each database.
3. Optional (project specific) tables and fields – These are tables and fields that are needed for a specific project or protocol, but will not likely be used in other databases or in integration.

The NRDT has been extensively reviewed by the I&M Program data managers and has been applied to some of the biological inventory projects in the Program. It has not, however, been applied extensively to monitoring. Although the fundamentals of this database structure will be applied to GLKN databases, its applicability towards Vital Signs monitoring may require specific mandatory tables to be reconsidered depending on the sampling methods, information collected, and potential for integration with the GLKN geodatabase GIS structures.

4.6.6. Documenting and Integrating Databases

Database documentation is more fully described in Chapter 7 (Data Documentation).

Each GLKN project will rely on a modular, stand-alone project database/geodatabase that is developed based on these shared design standards. The GLKN will develop master look-up tables, residing in the enterprise SQL RDMS, which can be used in multiple project databases to allow databases across projects and across disciplines to remain as consistent as possible. Standard attributes and values likely to be contained in these master look-up tables include park codes, watershed codes, species identifiers, and vegetation or land type classifications. To the extent possible, GLKN will work with WASO and other I&M networks to develop and utilize shared look-up tables for data consistency and inter-operability at a higher level within the I&M Program.

Credits

This chapter was adapted from concepts and material developed by John Boetsch (North Coast and Cascades Network), Dorothy Mortenson (Southwest Alaska Network), and Gordon Dicus (Pacific Island Network).

5. DATA ACQUISITION AND PROCESSING

The National Park Service's Inventory and Monitoring Program, in support of NPS' Natural Resources Challenge, is responsible for acquiring information needed by park managers to properly manage and maintain the natural resources of their parks. To successfully accomplish this task, the GLKN collects information from multiple sources and processes it to meet national and Network standards. This chapter describes the steps involved with acquiring data as well as the initial stages of data processing.

Processes used for data acquisition are as follows:

- Data discovery
- Data harvesting
- Data collection for field studies
- Data collection for remote sensing
- Data compilation, processing, and integration

National Standards

Executive Order 12906, Section 3 (d) states that each agency must adopt internal procedures to ensure that the agency accesses the National Geospatial Data Clearinghouse before it expends federal funds to collect or produce new geospatial data to determine whether the information has already been collected by others or whether cooperative efforts to obtain the data are possible.

Network Standards

The GLKN will use the National Geospatial Data Clearinghouse for searching existing geospatial data, as stated above. In addition, GLKN will use the internet, agency contacts, and other means to find other spatial and non-spatial data that may benefit the Network's program.

5.1. Data Sources

There are two general classifications for the types of data handled by the I&M Program:

1. Programmatic data – any data produced from projects that are initiated (funded) by the I&M Program or projects that in some way involve the I&M Program.
2. Non-programmatic data – includes data collected from NPS sources and data produced by external non-NPS sources.
    - Non-programmatic NPS data – any data produced by the NPS that did not involve the inventory and monitoring program, such as park visitor use information.
    - Non-programmatic external data – any data produced by agencies or institutions other than the National Park Service, such as the National Oceanic and Atmospheric Administration (NOAA) or a state Department of Natural Resources.

The Network will likely use a combination of both programmatic and non-programmatic data to meet its monitoring goals.

5.2. Data Discovery

Data discovery or data mining is the process of searching for existing data or information that may be useful to the I&M Program mission and that is related to the natural resources of the Network parks. This is a continual process that begins with the collection of background information and data to assist in the development of project methodologies and designs. The process involves reviewing many different sources for varying types of information. Although data mining is an important part of any I&M project, some Vital Signs monitoring projects may depend largely on data collected by other agencies or institutions and harvested through Network data mining endeavors. Many of the data sources listed below are accessible via the internet, but some require visiting local archives, research or academic institutions, museums, or local parks in order to search reference material.

Bibliographic/Literature

- National NPS databases (e.g., NatureBib)
- Online literature databases (e.g., First Search or Biosis)
- Park archives through ANCS+

Geographic Data

- National centralized GIS data from NPS Metadata and Data Store
- Federal Geographic Data Clearinghouse(s)
- Local, state, and federal government offices
- Regional GIS specialists

Biological/Natural Resources Data

- NPSpecies
- Voucher collections (museums, parks, universities)
- Network parks' records and documents
- Local, state, and federal government offices

Relevant information collected during a data discovery process is maintained at the Network either electronically or in hard copy format, depending on how it was collected. Any data collected during data discovery should be accompanied by as much documentation (metadata) as possible. Geographic datasets collected during this process are documented with FGDC-compliant metadata. In some cases, datasets will be documented using the metadata tools described in Chapter 7.

5.3. Data Harvesting

The process of harvesting data from other sources should be standardized as much as possible. Prior to harvesting data, the following questions should be considered:

- Does the data source organization know of the I&M Program and its data needs?
- Does a Memorandum of Understanding (MOU) need to be in place?

- What is the contingency if this data source is no longer available?
- Can downloads or requests for data be consistently exported/imported and scheduled?
- How will the downloaded data be stored and integrated into the I&M Program?
- How should errors be addressed?
- Is the data source organization interested in integrating I&M Program data where appropriate?
- Is the documentation adequate?

For Vital Signs monitored through data harvesting, these questions should be addressed in their data management protocols.

Data harvested from another agency may not exactly match the needs of the GLKN. Boundaries used for summary data often do not coincide with park boundaries, or specific parameters may not be currently collected. In addition, data may not meet minimum GLKN accuracy standards. Project managers are encouraged to work with these agencies to incorporate GLKN needs and standards as much as possible. Money can be saved and more efficient monitoring can be accomplished by improving existing systems.

5.4. Data Collection for Field Studies

Biological inventories and monitoring projects are the most common examples of field studies conducted by the GLKN I&M Program. The project manager is responsible for ensuring that data collection, data entry, verification, storage, and archiving for all field projects are consistent with GLKN standards. Project managers will make use of GLKN general standard operating procedures (SOPs) and guidance documents as applicable. The data manager will work closely with the principal investigator and Network staff to develop protocol-specific SOPs for the collection, storage, and maintenance of project data. This may range from detailing the proper usage of data entry forms or databases to outlining calibration procedures for automated data loggers.

Some of the tools that can be used to collect field data are listed below. This list serves as a guide for consideration as the Vital Signs protocols are developed. Details on how these tools may apply to protocol-specific SOPs will be contained in the individual protocol documentation.

- **Field Forms** – These are the most common method of recording field data. Formatted, project-specific data sheets are recommended as opposed to a field notebook. Field notebooks are important for entering additional notes and observations. Acid-free paper should be used to prevent fading and subsequent data loss. Some circumstances may warrant the use of paper and writing implements that can withstand moisture, dust, and other extreme environmental conditions (e.g., Rite in the Rain® paper). Standardized data sheets that identify the pieces of information to be recorded and forms that reflect the design of the computer data entry interface help ensure that all relevant information is recorded and that subsequent data entry errors are minimized. Data sheets should contain as much basic, preprinted project information as possible and should provide sufficient space to record relevant metadata (e.g., date, collectors, weather

conditions). They should clearly specify all required information, using examples where needed, to ensure that the proper data are recorded. Data recorders should adhere to the following guidelines:

- All information added to the data sheet must be printed and clearly legible.
- If alterations to the information are necessary, the original information should be crossed out with a single line and the new information written next to the original entry. Information should never be erased and old information should not be overwritten.

Upon return from the field, all original data sheets should be copied and checked for legibility and completeness (i.e., no data cut off at the edges). The original data sheets should be used for data entry and then they should be stored in the fireproof cabinet. Copies of original data sheets should be archived offsite.

- **Tape Recorders** – Handheld micro-cassette tape recorders are useful for recording field data. Recorded observations are subsequently transcribed to paper or directly entered into computer files. As with other technological solutions, there are drawbacks (e.g., battery and tape maintenance, low environmental tolerance, risk of failure). However, if a single data collector is in the field, tape recorders can provide an easily-operated, high quality, efficient method of collecting data. All audio tapes used for recording field data should be labeled appropriately (e.g., date, site, project) and stored in the fireproof cabinet. Magnetic audio cassettes degrade over time and are a medium that is quickly becoming outdated and obsolete. If this method is used to collect field data, efforts should be made to transfer the audio data to a more permanent audio format such as CDs or MP3 files. Preference for recording aural measurements or observations in audio should be to digital, non-volatile memory due to the reduced risk of data loss or corruption.
- **Cameras** – Photographs provide an excellent visual record of field visits. Cameras are useful for capturing photo point records of long-term study sites. They also serve well for automated data collection by remotely recording visual information using continuously recording webcams or event-triggered cameras.
- **Field Computers** – Data collection and data entry efficiency can be increased by eliminating the need for paper field forms. Data can be downloaded directly from field computers to office desktop computers, thereby eliminating the data entry step. This results in less chance for error because QA/QC checks can be built into the field database. The use of field computers can be inefficient if copious amounts of notes or comments need to be recorded in the field.

  Field computers, however, are vulnerable to environmental conditions such as heat, dust, and moisture. When handheld computers are used for data entry in the field, the data should be downloaded daily to avoid potential loss of information. No field computer should be used unless it is equipped with a removable flash memory card to store field data in case of a catastrophic failure of the field unit. Batteries should be checked prior to a data collection trip, and they should be charged at the end of every field day. Also, in case the unit becomes inoperable in

the field, printed data sheets should always accompany field teams on data collection trips.

- **Palm-top computers (PDAs)** – The small size and relatively low cost of these devices make them attractive options for entering field data. They are good for small projects but not powerful enough for large, data-intensive field projects. These devices can be ruggedized fairly easily and at a relatively low cost. Most run either Windows CE or Palm operating systems and may require additional processing or programming to transfer the database structure to the field units.
- **Tablet PCs** have the same properties as most laptops and provide the user with the convenience of a touch screen interface. They are bulkier, more expensive, and harder to ruggedize than PDAs, but are more powerful as well. Tablet PCs are good for field projects that are very data intensive. Because these units run Windows XP (Tablet Edition), the project database can be directly transferred from desktop units to field units without additional programming steps.

- **Automated Data Loggers** are mainly used to collect ambient information such as air or water quality information. In addition to environmental sensors, this category includes auto-triggered event recorders or cameras, automated RFID readers, and telemetry receivers. These devices can be calibrated and programmed to automatically record data and store them for later download to a computer, thereby eliminating the possibility of manual data entry errors. Automated units must be properly calibrated and maintained and field crews must be properly trained. Calibration and maintenance procedures of the units must be documented in SOPs. Standard operating procedures are currently in development and will be distributed to all field crews using these devices.
  - **Permanently deployed devices** – The high initial cost of permanent devices may be offset by reduced labor costs if fewer field visits are required. Data from these devices must be retrieved and batteries must be changed on a regular basis. The intervals for each activity should be defined in the protocol.
  - **Portable hand-held devices** – These units are deployed for sampling only during site visits. They are generally less expensive than units that are permanently deployed in the field. Prior to and following field visits, device components should be inventoried to ensure that all necessary equipment is accounted for.
- **GPS Units** – Two types of GPS units are often used during field work in GLKN parks to collect location information.
  - **Recreational Grade GPS Receivers (e.g., Garmin)** are good for collecting general position information. They are not suitable for obtaining high-accuracy location information, but are suited for navigating to preexisting marked locations or for recording incidental location information.
  - **Mapping Grade GPS Receivers (e.g., Trimble)** are good for collecting highly accurate (submeter) location information. They are usually employed

with well developed forms or a database application that ties in with a desktop processing program and GIS standards. The increased cost of these units may be outweighed by their additional accuracy and simplified data flow capabilities.

5.5. Data Collection for Remote Sensing

Remote sensing technologies can be a powerful tool for characterizing and analyzing landscape data, as well as capturing data within areas of low accessibility. Considerations for selecting remote sensing imagery are as follows:

- Accuracy and spatial grain or resolution needed
- Extent of imagery required
- Frequency of measurement
- Cost of acquisition and processing
- Licensing required for public use
- Need for ortho-rectification

Each remote sensing product is unique. It is imperative that the user fully understands the product being used. Products should be accompanied by well-documented metadata. Imagery that has not been rectified cannot be used to measure distance or area, but may provide a low cost and timely overview of the landscape and be valuable in the field. For protocols that use remote sensing data, project managers should consult with the Network GIS Specialists and consider the trade-off between accuracy and costs among different imagery sources.

Although cost will be a primary deciding factor, consistency among park units should also be considered. Ideally, all imagery will be received in electronic and geo-referenced format.

The Network has developed a substantial GIS catalog, which includes base cartographic layers for the parks, (e.g., digital orthophotos and topographic maps in standard quadrangle tiles), natural resource theme layers, and satellite imagery (Appendix J). Recent, high-resolution orthophotos and satellite imagery have been acquired for several of the parks. As individual Vital Signs monitoring protocols identify additional remote-sensing needs, GLKN will pursue cost-efficient options to acquire imagery or data layers in suitable formats.

The Network is currently developing a long range plan for the acquisition of remote sensing products for the parks. The first consideration is to determine what questions can be answered with remotely sensed data, whether the need is to identify broad land cover classes, or to identify roads, buildings, trails, and small wetlands. The main considerations for choosing the appropriate remote sensing platform have to do with spatial grain (ground pixel resolution), spectral bands, extent of imagery (area of interest), cost, and timing of acquisition.

Because most of the Great Lakes parks are relatively small, ranging from 470 acres to 150,000 acres, high resolution data sources may be cost-effective. Whereas the cost of high resolution satellite imagery has become quite competitive, airborne sensors continue to be preferred for a number of reasons. The broad window for timing of acquisition,

cloud cover or haze allowance, and current spatial grain no better than 0.6 meter continue to be problematic with high-resolution satellite imagery. Recently developed digital mapping cameras now provide a wide spectral range, simultaneously capturing color infra-red, natural color, and black and white spectral ranges, with spatial grain as fine as 0.15 meter. Acquisition of high resolution terrain and canopy information using LiDAR (Light Detection and Ranging) is also becoming more common and affordable.

The Network will continue to pursue appropriate technologies to address the differing needs for land cover information while recognizing that these will continue to evolve rapidly. We will also continue to pursue cooperation with other agencies and organizations to find cost-effective solutions for remote sensing products.

5.6. Changes to Data Collection Procedures

Changes to data collection procedures will be based on valid reasons for altering the methodologies. Most issues will be identified while planning, designing, and testing the protocols, and changes will be implemented prior to collection of field data. Unforeseen issues that arise after data collection has begun may require revision of procedures or protocols.

Improvements in technology may also require changes to procedures. Significant changes to protocols must be approved by the project manager, the Network coordinator, and the data manager. The Network coordinator will determine if proposed changes to protocol procedures require additional peer review before the changes are accepted and implemented.

Changes to protocols and associated data collection procedures may also occur as a result of scheduled Program reviews. During the review, data may be analyzed to determine if the current protocol is meeting stated objectives. If a protocol has not achieved the desired results then changes should be recommended. All changes must be carefully documented within the SOPs and any associated databases.

5.7. Data Compilation, Processing, and Integration

All data, whether collected in the field by the Network or downloaded from another organization, will need to be compiled and processed. This compilation should follow data management guidelines developed for the specific data type (see Appendix B for GLKN Project Deliverables guidance document). For example, tabular data will need to be entered or imported into a pre-approved Microsoft Access database or geodatabase. Quality assurance and quality control (QA/QC) procedures are more fully described in Chapter 7.

After compilation and QA/QC is completed for data from a single field season or at a project milestone period, these data should be integrated with master datasets. Some extraction of information may be required. For example, species information may need to be extracted from various project databases for the purpose of updating the master NPSpecies database.

Credits

This chapter was adapted from material developed by Geoffrey Sanders (National Capital Region Network).

6. QUALITY ASSURANCE (QA) AND QUALITY CONTROL (QC)

The Great Lakes Network must have confidence in the data it uses and provides to others. Analyses to detect trends or patterns in ecosystem processes require high-quality, well-documented data. Data of inconsistent or poor quality can result in loss of sensitivity and incorrect interpretations and conclusions. The GLKN monitoring program will bring together analysis from widely-varying projects aimed at detecting changes in the environment over time. The inherent complexity in this endeavor demands that stringent and consistent quality assurance and control measures be applied throughout the program.

Quality assurance involves planning to obtain the highest possible data quality, while quality control consists of monitoring the system or appraising the product after the product is developed. The Network will establish and document protocols in order to identify and reduce error at all stages in the data lifecycle. These stages include project planning and database design, data collection, data entry, verification and validation (certification), documentation (including data quality and sensitivity review), and archiving (Figure 6.1). The final stage in the data life cycle is dissemination and integration. Detailed QA/QC procedures for these stages will be included in each protocol initiated by the GLKN. This chapter of the data management plan presents more broadly based procedures and policies that govern specific operations within a GLKN project. Figure 6.2 illustrates selected QA/QC procedures relative to the amount of planning and quality control necessary to have confidence in the data.

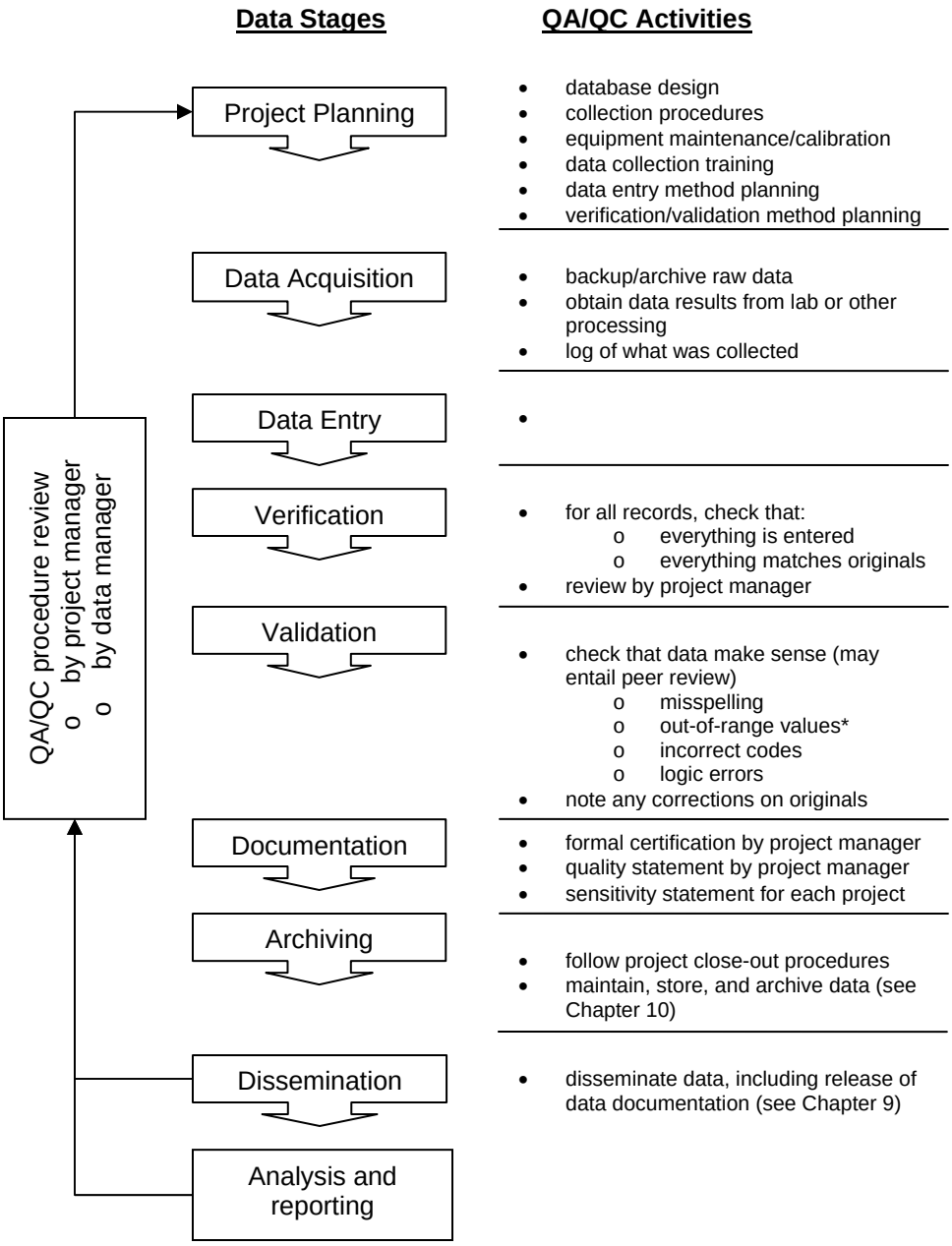
Quality assurance and quality control cannot be considered without specifying certain roles and responsibilities native to such procedures. The duties outlined in this chapter are inherited from those listed in Chapter 2 and conform to their letter and spirit. These duties should be followed in light of those given in Chapter 2.

A Word on Data Quality Expectations

Although a data set containing no errors would be ideal, the cost of attaining 95%-100% accuracy may outweigh the benefit. Therefore, we consider at least two factors when setting data quality expectations:

- frequency of incorrect data fields or records
- significance of error within a data field

We are more likely to detect an error when we work with clearly documented data sets and understand what a ‘significant’ error is within *that* data set. The significance of an error can vary with data sets and depends on where it occurs. For example, a two-digit number off by one decimal place is a significant error. A six-digit number, with the sixth digit off by one decimal place, is not a significant error. But one incorrect digit in a six-digit species number could indicate a different species. That is a significant error.



* CAUTION! Care must be exercised when culling 'out-of-bounds' data. See text for discussion.

Figure 6.1. General course of data and associated QA/QC procedures. Quality control with regards to data analysis is specific to each project and addressed in appropriate standard operating procedures.

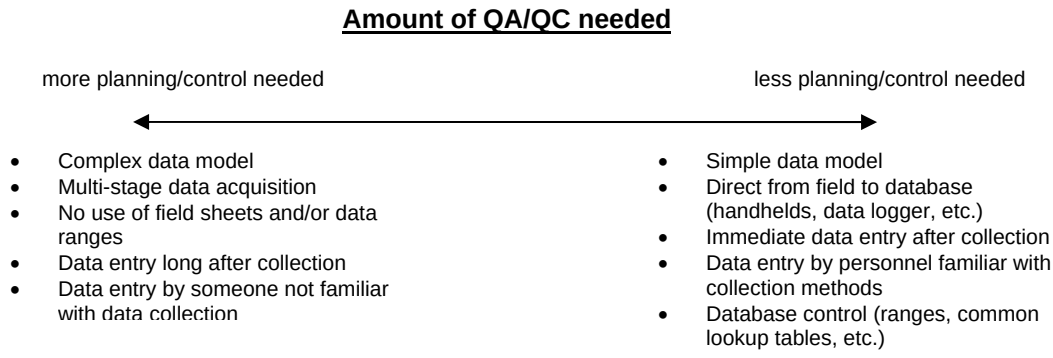


Figure 6.2. Some common data management elements affecting the amount of QA/QC needed. Planning and training for data collection (QA) and entry is always a premium.

6.1 National Park Service Mandate for Quality

Director's Order #11B, which was issued by the NPS in 2002, "Ensuring Quality of Information Disseminated by the National Park Service," defines *'quality'* as incorporating three key components–*objectivity*, *utility*, and *integrity*.

Objectivity consists of: 1) *presentation,* which focuses on whether disseminated information is being presented in a proper context, in an accurate, clear, complete, and unbiased manner; and 2) *substance,* which focuses on the accuracy, usability, and reliability of the information.

Utility refers to the usefulness of the information to its intended users, from the perspectives of both the Network and the general public.

Integrity refers to the *soundness* of the data or the confidence one has in the data. Integrity is integrally related to objectivity; however, it is possible to have subjective data of high integrity. The integrity of data is also related to data security. Data must, for example, be protected from unauthorized access or revision to ensure that the information is not compromised through corruption or falsification.

Director's Order #11B also specifies that information must be based on reliable data sources, which are accurate, timely, and representative of the most current information available. These standards apply not only to NPS-generated information, but also to information provided by other parties to the NPS if the NPS disseminates or relies upon this information.

High quality data and information are mandated by directives and orders, and they are vital to the credibility and success of the I&M Program. In a statement issued in 2001, Abby Miller of the Natural Resource Stewardship and Science Division, said "data need to meet national-level quality standards and need to be accessible to be used for wise and defensible decision-making at all levels. Data need to be able to be shared and aggregated with data from other parks and from adjacent lands to support landscape-level and national planning and decision-making."

6.2. Quality Assurance and Control Duties

Producing and maintaining high quality data is the responsibility of everyone involved with the handling of project data. While Chapter 2 specifies data management roles and responsibilities, selected QA/QC duties are emphasized here.

Project managers must:

- be aware of quality protocols and convey their importance to technicians and field crews
- ensure compliance with the protocols
- plan for and ensure proper execution of data verification and validation
- review all final reports and information products

Technicians must follow established protocols for data collection, data entry, and verification established in the project SOPs.

The data manager is responsible for:

- developing Network-wide protocols and SOPs to ensure data quality
- making project managers, technicians, etc., aware of the established procedures and enforcing adherence to them
- evaluating the quality of all data and information against NPS standards before dissemination outside the Network
- performing periodic data audits and quality control checks to monitor and improve quality control operations

6.3. Data Quality Goals and Objectives

The Network must ensure that a project produces data of the right type, quality, and quantity to meet project objectives. Quality criteria should be proportionate to project-specific objectives and these criteria should indicate the level of quality that is acceptable. Project subjects and goals will drive data quality needs and control the kinds of analysis and summarization that may be defensibly applied.

The most effective mechanism for ensuring that a project produces data of the right type, quality, and quantity is to provide procedures and guidelines to assist the researcher in accurate data collection, entry, and validation. As part of data management operations, the GLKN will develop a comprehensive set of SOPs and quality assurance guidelines to be used in the collection, entry, validation, verification, and use of data. This will likely include development of new field-computer-based methods for acquiring data as well as custom programming of database interfaces to both protect the data and control how the data are viewed or summarized.

6.4. General Operations

6.4.1. Version Control and File Naming Standards

Version control is the process of managing copies of changing files over the course of a project; file naming standards are critical for effective version control. Because of the

scope and quantity of files being consolidated into one place, it is critical that computer files be given names that will uniquely identify them and indicate their content, even over the course of time.

The following conventions apply to all GLKN file names:

- No spaces or special characters within the name
- Include date for version control, in yyyymmdd format (e.g., 20051104 for Nov. 4, 2005)
- Use underscore as delimiters

File names for final products will begin with "GLKN" and contain a brief but clear explanation of the file content followed by an eight digit date and the extension. Most file names will require an indication of the QA/QC status of the information as a name element. Examples of QA/QC status include "draft", "raw", "verify", and "valid". Each of these file name elements (excepting the extension) is preceded by an underscore ("_"). Additional underscores may be added for clarity. As an example, "GLKN_ Data_Mngmt_Plan_draft_20051128.doc" indicates a draft version of the data management plan dated Nov. 28, 2005.

6.4.2. Version Control

Before making major changes to a file, a copy of the file with the appropriate version control (in this case the file name) should be made. As indicated above, files are stored with the appropriate eight digit date which serves as version control. This allows changes to be tracked over time and facilitates collaboration between multiple personnel working on common files. With proper controls and communication, versioning ensures that only the most current version is used in any analysis. Including the date, formatted as yyyymmdd, in the file name provides logical version control.

6.4.3. External Data

Several of the GLKN Vital Signs involve laboratory analysis conducted via contract with certified laboratories. All data for a given Vital Sign, including lab results, will be housed in a distinct relational database accessed via custom applications built in Microsoft Access and in corresponding SQL Server project tables. Laboratories that will be providing data for a given Vital Sign will be supplied with a copy of the application so that data may be entered in a format consistent with the rest of the data for that Vital Sign. While most professional laboratories exercise their own QA/QC procedures, results received by a project manager are subject to the same QA/QC measures exacted on other project data.

6.5. Project Planning and Data Design (Quality Assurance)

The methods and array of information a given project chooses to use and gather are integral to data quality. The GLKN holds as policy that techniques and procedures chosen for a project will maximize data quality. Quality assurance will be achieved by stipulating that:

- Common lookup tables are created for values of parameters recorded in an identical field for more than one project (such as common weather metrics)

- Each project has SOPs for core data management (collection, entry, verification, etc.)
- Each project use, at a minimum, standard field sheets for data collection (see Chapter 5)
- Database data entry computer applications resemble field sheets
- Automated error checking features be included in database applications
- Database application design will maximize the use of auto-fill, auto-correct, value range limits, pick lists, and other constraints specific to projects
- Database applications will include a means to track errors reported on the data after dissemination
- Database maintenance logs will be maintained for each GLKN database and housed in association with database files

6.5.1. More on Database Design - Record-level Tracking

As a standard part of database design, the GLKN will build into database tables, fields that track at the record-level who entered the data, precise entry time and the protocol version under which the data were collected. The benefits to overall data integrity outweigh any inconveniences this 'overhead' data may cause due to factors such as increased database file size.

6.5.2. More on Lookup Tables

As noted in Section 2.1, the GLKN will utilize to the fullest extent possible common lookup tables for variables recorded by multiple projects. Examples include weather variables (such as precipitation intensity, wind speed, etc.), standard equipment and settings (e.g., GPS models and datums) and possible field personnel. Section 6.9 addresses database programming used for data validation.

6.5.3. More on Project SOPs

Each Vital Sign protocol will include SOPs that address core data management practices with quality control in mind. These may include:

- Field crew training (addressing both data collection and entry)
- Standardized data sheets
- Use of handheld computers
- Equipment maintenance and calibration
- Procedures for handling data (including specimens) in the field
- Data backup, entry, verification, and validation

6.6. Data Collection

Chapter 5 addressed the general acquisition and initial handling of data. Attention to detail during data collection is crucial to overall data quality. The GLKN adopts the following precepts regarding data collection that affect data quality. Changes to data collection protocols are addressed in Chapter 5.

- At a minimum, data will be collected on formatted, project-specific data sheets that reflect the overall data design for the project and maximize limitations on values that may

be recorded for different parameters. Sheets will be designed to minimize the amount of writing necessary to effectively record observations. Data sheets will be printed on a stock appropriate to field conditions (generally Rite in the Rain® paper).

- Any project using field equipment will include an SOP for calibration (including the timing of calibration) and maintenance. Such SOPs will specify establishment of an equipment maintenance log.
- Edits on field sheets while in the field will be done by drawing a single line through the information to be changed and adding any replacement information in clear writing next to the original entry.
- Completed field forms will be proofed for errors each day in the field.
- Wherever possible and appropriate, data loggers or field-based computers will be used to collect data. The use of handheld computers requires separate SOPs to describe their use and will include direction for daily review and back-up of digital data.

6.6.1. More on Field Sheets

Standardized data sheets that identify the pieces of information to be recorded and forms that reflect the design of the computer data entry interface will help ensure that all relevant information is recorded and subsequent data entry errors are minimized. Data sheets should contain as much basic preprinted project information as possible and sufficient space for recording relevant metadata such as date, collectors, weather conditions, etc. They should clearly specify all required information, using examples where needed to ensure that the proper data are recorded. Data collectors should adhere to the following guidelines:

- All information added to the data sheet must be printed and clearly legible.
- If alterations to the information are necessary, the original information should be crossed out with a single line and the new information written next to the original entry. Information should never be erased and old information should not be overwritten.
- Upon return from the field, copies of all original data sheets should be made and checked for legibility and completeness (i.e., no data cut off at the edges). The copies of the data sheets will be stored as specified in the protocol SOP, and the original data sheets will be used for data entry.

6.7. Data Entry or Import

'Data entry' is the initial set of operations where raw data are transferred to database tables using a computerized form. When data are gathered or stored digitally in the field (e.g., on a data logger), data entry consists of transferring the data (downloading) to a file in an office computer where they can be further manipulated. The goal of data entry is to transcribe field observations into a computer database with 100% accuracy. In other words, that which is recorded in the field should be entered exactly in the database. Subsequent data verification is conducted to ensure that raw data matches entered data. Following verification, data validation may result in changes *to the entered data.* Data entry is a separate operation from data validation and care must be taken to not impose validation (beyond that automatically imposed by programming rules in a database) during data entry.

The GLKN adopts the following precepts affecting data quality with regards to data entry:

- Data will be entered as soon as reasonably possible after collection.
- Data entry will be done by someone familiar with data collection. The project manager (with assistance from the data manager if needed) must ensure that data entry staff are familiar with the database software, database structure, and any standard codes for data entry used by the Network. At a minimum, data entry technicians should know how to open a data entry form, create a new record, and exit the database properly. They must learn how to commit both a 'field' entry and a 'complete record' entry and to correct mistakes made while typing.
- If feasible, data will be entered by two qualified persons; one person will read the observations and the other will enter the data.
- Data will be entered into pre-designed database forms that resemble field sheets to maximize error control. Whenever possible, data should not be entered into spreadsheets.
- To the extent possible, data entry will be automated. This may simply entail downloading data from field-based computers but may include the application of new technology to allow for machine-driven data entry.

6.8. Data Verification (Quality Control Part 1)

We appraise data quality by applying verification and validation procedures as part of the quality control process. These procedures are more successful when preceded by effective quality assurance practices (planning). *Data verification* checks that the digitized data match the source data, while *data validation* checks that the data make sense. It is essential that we validate all data as truthful and do not misrepresent the circumstances and limitations of their collection. Failure to follow SOPs for data entry, validation, and verification will render a data set suspect. Although data entry and data verification can be handled by personnel who are less familiar with the data, validation requires in-depth knowledge about the data.

The GLKN adopts the following precepts regarding data verification:

- Options for verification methods are given below. Project managers are responsible for specifying in the project protocol one or more of these methods. At the discretion of the project manager, additional verification methods may be applied.
- The project manager is responsible for proper execution of data verification (see Chapter 2).
- Data verification is carried out by staff sanctioned by the project manager, who should be familiar with data collection and entry conditions and techniques.
- All records (100%) will be verified against original source data.
- A subset of randomly selected records, (10% of records), will be reviewed after initial verification by the project manager. If errors are found, the entire data set should be verified again.

- A record of each dataset's verification process, including number of verification iterations and results, will be prepared by the project manager as part of formal metadata generation (see Chapter 7 for more details).

6.8.1. Methods for Data Verification

Each of the following methods has a direct correlation between effectiveness and effort. The methods that eliminate the most errors can be very time consuming while the simplest and cheapest methods will not be as efficient at detecting errors.

1) *Visual review at data entry.* The data entry technician verifies each record after input. Recorded values in the database are compared with the original values from the hard copy. Identified errors are immediately corrected. This method is the least complicated since it requires no additional personnel or software. Its reliability depends entirely upon the person keying data and thus, is probably the least reliable data verification method.
2) *Visual review after data entry.* All records are printed upon completion of data entry. The values on the printout are compared with the original values from the hard copy. Errors are marked and corrected in a timely manner. If only one technician is available to perform the review, the method's reliability increases if someone other than the person keying the data performs the review. Alternatively, two technicians can perform this review. One technician reads the original data sheets (the reader), and the second reads the same data on the printout (the checker).
3) *Duplicate data entry.* The data entry technician completes all data entry, as normal. Random records are selected (every **n**th record) and entered into an empty replica of the permanent database, preferably by someone other than the person keying the permanent data. A database query is then used to automatically compare the duplicate records from the two data sets and report any mismatches. Disparities are manually reviewed and correction applied if necessary. This method adds the overhead of retyping the selected records, as well as the creation of a comparison query. However, it becomes increasingly successful as the value of **n** decreases. Professional data entry services frequently use this method.

6.8.2. Supplementary Methods

Data verification can be improved by calculating summary statistics and by identifying duplicate or omitted records. For example, we can view the number of known constant elements, such as the number of sampling sites, plots per site, or dates per sample. Posing questions in different ways and looking for differences in the answers can reveal errors. The more checks we devise to test the completeness of the data, the greater our confidence that we have completely verified the data.

6.9. Data Validation (Quality Control Part 2)

Validation is the process of reviewing computerized data for range and logic errors and may accompany data verification *only* if the operator has comprehensive knowledge of the data and subject. More often, validation is a separate operation carried out *after*

verification by a project specialist who can identify generic and specific errors in particular data types.

General step-by-step instructions are not possible for data validation because each data set has unique measurement ranges, sampling precision, and accuracy. Nevertheless, validation is a critically important step in the certification of the data and a required component of any GLKN project protocol. Invalid data commonly consist of slightly misspelled species names or site codes, the wrong date, or out-of-range errors in parameters with well defined limits (e.g., pH). But more interesting and often puzzling errors are detected as unreasonable metrics (e.g., stream temperature of 70°C) or impossible associations (e.g., a tree 2 feet in diameter and only 3 feet high). We call these types of erroneous data *logic errors* because using them produces illogical (and incorrect) results. The discovery of logic errors has direct, positive consequences for data quality and provides important feedback to the methods and data forms used in the field. Histograms, line plots, and basic statistics can reveal possible logic and range errors.

The GLKN adopts the following precepts regarding data validation:

- Each GLKN project protocol will address a process for data validation that includes at least one of the methods outlined in the Network data management plan. That process will adhere to the precepts given in the plan.
- Corrections or deletions as a result of data validation require notations in the original paper field records about how and why the data were changed, with the editor's initials.
- Modifications of the field data should be clear and concise while preserving the original data entries or notes (i.e., no erasing!).
- Validation efforts should also include a check for the completeness of a data set since field sheets or other sources of data could easily be overlooked.
- The GLKN will maximize the use of automated routines and/or data summary and visualization such as histograms, line plots, and basic statistics to reveal possible logic and range errors.
- The GLKN will maximize the use of database programming to control data entry. In general, this will be achieved via the use of lookup tables, but may be accomplished by field-type design in a data base (such as "yes/no" field-types).

6.9.1. Methods for Data Validation

Specific procedures for data validation depend upon the vital sign being monitored (project subject) and will be included in the project protocols. The following general methods can be used to validate data:

1) *Data entry application programming.* Certain components of data validation are built into data entry forms. This method is essentially part of database design and is discussed earlier in this chapter. Not all fields, however, have appropriate ranges that are known in advance, so knowledge of what are reasonable data and a separate, interactive validation stage are important.

   Caution must be exercised when using lookup tables to constrain variable values. Values occurring outside the range set by a lookup table (established during database design) may not always be invalid. As part of data validation

procedures, the project manager is responsible for correct use of lookup tables or other automated value range control.

2) *Outlier Detection.* According to Edwards (2000), "the term outlier is not (and should not be) formally defined. An outlier is simply an unusually extreme value for a variable, given the statistical model in use." Any data set will undoubtedly contain some extreme values, so the meaning of 'unusually extreme' is subjective. The challenge in detecting outliers is in deciding how unusual a value must be before it can (with confidence) be considered 'unusually' extreme.

   Data quality assurance procedures should not try to *eliminate* outliers. Extreme values naturally occur in many ecological phenomena; eliminating these values simply because they are extreme is equivalent to pretending the phenomenon is 'well-behaved' when it is not. Eliminating data contamination is a better way to explain this quality assurance goal. If contamination is not detected during data collection, it is usually only detected later if an outlying data value results. When an outlier is detected, an attempt should be made to determine if some contamination is responsible.

   Database, graphic, and statistical tools can be used for ad-hoc queries and displays of the data to detect outliers. Some of these outlying values may appear unusual but prove to be quite valid after confirmation. Noting correct but unusual values in documentation of the data set saves other users from checking the same unusual values.

3) *Other exploratory data analyses.* Palmer and Landis (2002) suggest that in some cases, calculations for assessments of precision, bias, representativeness, completeness, and comparability may be applicable and that for certain types of measurements, evaluation of a detection limit may also be warranted (these authors provide examples of procedures that may be applicable). Normal probability plots, Grubb's test, and simple and multiple linear regression techniques may also be used (Edwards 2000).

6.10. Data Quality Review and Communication

The National Park Service requires QA/QC review prior to communicating or disseminating data and information. Only data and information that adhere to NPS quality standards may be released (see earlier treatment in this Chapter). Data and information disseminated to the public must be approved by the appropriate reviewing officials and programs. Documentation of the QA/QC standards used in producing the information and that substantiate the quality of the information must be formally certified and distributed with related data and information. Also, mechanisms must be in place for receiving and addressing comments or complaints pertaining to the quality of data. Chapter 9 addresses the dissemination of GLKN data and information.

6.10.1. Monitoring Conformance to Plans and Standards

As part of the close-out and evaluation stage of each GLKN project, QA/QC procedures will be reviewed by the project manager and recommendations for change will be included in the annual report. Additionally, the GLKN data manager will review and revise the QA/QC procedures included in the data management plan as needed. While the

data management plan will be reviewed on a regularly scheduled plan (see Section 1.5 and Appendix B), QA/QC procedures are subject to revision as needed by the Network data manager.

The data manager will conduct periodic, random 'spot checks' of GLKN monitoring projects to ensure compliance with the data management plan and project protocol QA/QC procedures. The data manager will track and facilitate the correction of any deficiencies found during this process. The data manager will submit a report of findings to the project manager and the Network coordinator within a month of completing any review of QA/QC procedures. The project manager and Network data manager are responsible for ensuring that non-conformities in data management practices are corrected.

Periodic checks by the data manager to see if Network staff are adhering to the data quality procedures established in the Data Management Plan and protocols SOPs may include verification that:

- Data collection and reporting requirements are being met
- Data collection and reporting procedures are being followed
- Verification and validation procedures are being followed
- Data file structures and maintenance are clear, accurate, and according to plan
- Revision control of program documents and field sheets are adequate
- Calibration and maintenance procedures are being followed
- Seasonal and temporary staff have been trained in data management practice
- Metadata collection and construction for the program proceeds in a timely manner
- Data are being archived and catalogued appropriately for long term storage

6.10.2. Documenting and Communicating Quality

The final step in the QA/QC for a given dataset is the preparation of summary documentation that assesses the overall data quality. A statement of data quality will be composed by the project manager and incorporated into formal metadata as well as the GLKN primary data repository. Metadata for each dataset or database will also provide information on the specific QA/QC procedures applied and the results of the review. Typically, data quality information will be conveyed as part of FGDC-compliant metadata (see Chapter 7 for data documentation). Metadata and data will be available via both the GLKN website and the NPS NR-GIS & Metadata Data Store (see Chapter 9 for dissemination information).

Credits

This chapter was adapted from concepts and material developed by Debbie Angell (Sonoran Desert Network), Doug Wilder (Central Alaska Network), and Gordon Dicus (Pacific Island Network).

7. DATA DOCUMENTATION

The information needed to understand and use data is embedded in project documents such as study plans, reports, and protocols. Data design documents such as database descriptions and field sampling protocols are often critical to effective analysis of project data. Project tracking applications used by managers may also contain indispensable information about project data. Further, standard data handling such as backup, entry, verification, and validation may produce information that affects the way in which the data may be used.

Formal, standard metadata serves as a unifying document for all these sources of information about project data (Figure 7.1). While metadata are addressed in context throughout this plan, this chapter outlines the GLKN strategy for generating and handling metadata which serve as the principal documentation for data.

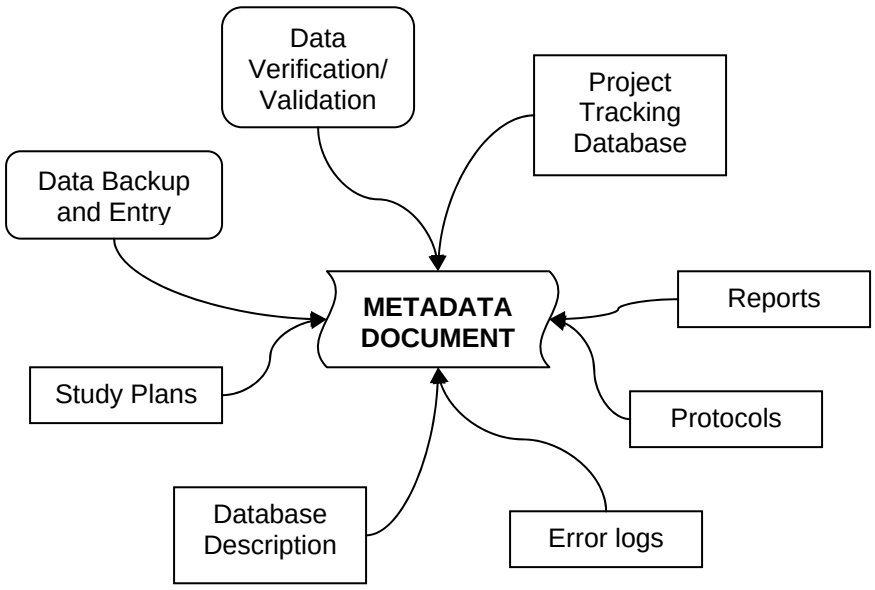


Figure 7.1. Project and data documentation. The metadata document contains all information pertaining to project data.

7.1. Mandate for Documentation

Executive Order 12906 mandates that federal agencies "...document all new geospatial data it collects or produces, either directly or indirectly..." using the Federal Geographic

Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM, *www.fgdc.gov/metadata/contstan.html*). In addition, EO 12906 directs agencies to plan for legacy data documentation and to provide metadata and data to the public.

The FGDC Biological Data Profile (*www.fgdc.gov/standards/status/sub5_2.html*) contains all the elements of the CSDGM and includes additional elements for describing biological data sets. Metadata created in compliance with the Biological Data Profile can be added to the National Biological Information Infrastructure (NBII, *www.nbii.gov/datainfo/metadata*) Clearinghouse. Although not a requirement, completion of the Biological Data Profile for appropriate data sets is recommended.

All GIS data layers must be documented according to applicable FGDC and NPS metadata standards. The NPS GIS Committee requires that all GIS data layers be described using FGDC standards and the NPS Metadata Profile (*nrdata.nps.gov/profiles/NPS_Profile.xml*).

While there are numerous tools available for developing metadata, the NPS Integrated Metadata System Plan (*science.nature.nps.gov/im/datamgmt/metaplan.htm*) recommends using one of three desktop applications: NPS Metadata Tools & Editor, NPS Dataset Catalog, and ESRI's ArcCatalog. Figure 7.2 illustrates the general NPS strategy for creating and distributing metadata.

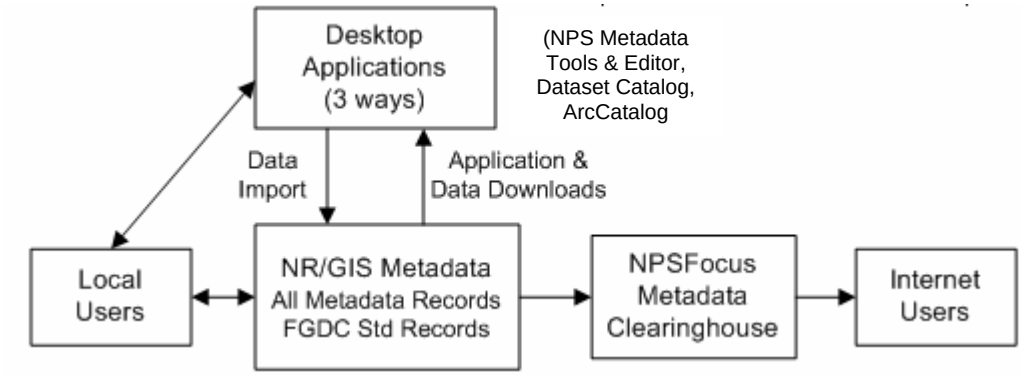


Figure 7.2. NPS Integrated Metadata System. The NR-GIS (NR/GIS) holds metadata records that are publicly accessible via the NPS Focus clearinghouse. Chapter 9 presents more information about NPS Focus.

7.2 Documentation Roles and Responsibilities

The project manager and Network data manager are responsible for documenting data, as shown in Table 7.1. See Chapter 2 for additional discussion of project roles and responsibilities concerning documentation.

Table 7.1. Data documentation responsibilities.

Project Manager	Data Manager
• Document project planning, execution, and progress in reports, correspondence, project-tracking database, etc. • Generate formal dataset metadata	• Facilitate metadata generation by providing tools and assistance in their use • Establish metadata dissemination protocols (cataloguing of metadata on accessible servers) • Format metadata for dissemination

The duties for the project manager, including direction of project personnel and general project management, are all aimed at producing a comprehensive metadata document.

The Network data manager is cast in the unwritten role of motivator of those with metadata generation responsibilities. The data manager ensures that metadata are available in formats suitable for distribution (see Metadata Formats below).

7.3. Documentation Process

7.3.1 Metadata Tools

As noted above, the NPS Integrated Metadata System Plan recommends three desktop applications for generating metadata:

1) NPS Metadata Tools & Editor (standalone or ArcCatalog plugin)
2) NPS Dataset Catalog (developed by the I&M Program)
3) ArcCatalog (commercial metadata tool)

Appendix H contains brief descriptions of these applications as well as a fourth tool, the Metadata Parser (mp) used to check metadata completeness and generate various metadata output formats.

The GLKN will utilize a combination of these three metadata tools (as well as the Meta Parser) to create metadata and related output.

The end goal, however, is a formal, comprehensive metadata document that meets federal (FGDC) and NPS standards. The tools and means of achieving this end product are numerous and changing. Project managers may use any tool deemed suitable for the project that meets program requirements.

7.3.2 Metadata Process

Figure 3.1 indicates the points within the standard project workflow for metadata development. Metadata development begins with project design and planning. The

GLKN data manager will establish a standard operating procedure for metadata generation and maintenance.

In general, a metadata document will be initiated using the NPS Metadata Tools & Editor, Dataset Catalog or ArcCatalog in the design and planning stage of a project. As the project progresses, this document will be augmented by the project manager to include relevant project details. Full development of metadata for project data will be completed after the dataset is certified by the project manager. Once metadata are complete, the Network data manager saves the document in one or more of four acceptable formats (Table 7.2) and parses the information into varying levels of information (Table 7.3). Metadata are saved on the primary GLKN server along with data and copied to the I&M Program NR-GIS metadata server (Figure 7.3). Hypertext links within the metadata document will point to accompanying datasets served on the primary GLKN server. In general, a single metadata document will apply to both raw and certified versions of the data. Section 7.9, below, further specifies that derived data (typically data resulting from the analysis of certified data) may require a separate metadata document. Metadata records will be stored with both hard copy and digital archive data.

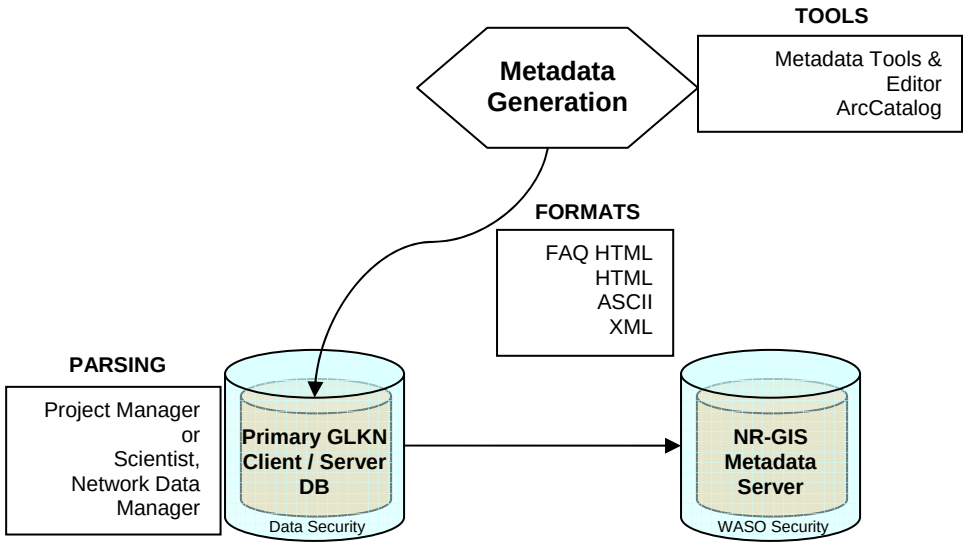


Figure 7.3. Metadata repositories for the GLKN. The GLKN server acts as primary storage of metadata. Metadata are available publicly via the NPSNR-GIS Metadata Data Store.

7.4. Metadata Formats

Table 7.2 lists the metadata file formats to be saved on the primary GLKN server.

Table 7.2. Metadata file formats.

Format	Purpose
HTML	Standard format for web browsers
FAQ HTML	Frequently Asked Questions version of metadata
ASCII	Standard text file
XML	Extensible Markup Language for application programming

These file formats will be generated using the Meta Parser (Appendix H) and will be viewable on the GLKN website. The XML format will be the preferred format for metadata exchange and archiving. The HTML file version will serve as the primary file for display on the website, and internal hyperlinks will allow for easy viewing, saving, and printing of the other formats.

7.5. Metadata Parsing

Metadata for primary server GLKN data will conform to FGDC guidelines and be parsed into three nesting levels of detail each designed with a specific audience in mind (Table 7.3).

Table 7.3. Metadata parsing strategy.

Metadata Parsing	Purpose
Level 1, or "Manager Level"	Overview of the dataset crafted to quickly impart the essential information needed to understand the product
Level 2, or "Scientist Level"	Additional details that allow for rapid scientific assessment of the product
Level 3, or "Full Metadata"	All components of supporting information such that the data may be confidently manipulated, analyzed, and synthesized

Parsed metadata will be viewable and printable via the GLKN website. The XML version of the metadata document will serve as the basis for parsing metadata.

7.6. Metadata Maintenance

The GLKN procedures for metadata generation and maintenance (see Section 7.3) will provide details on the mechanics of handling metadata. Generally, metadata will be created by the project manager, with assistance from the data manager, via an application interface tied to project data on the primary GLKN server. The Network data manager ensures that updated metadata are copied to the I&M Program NR-GIS Data Store.

7.7. Protocol Versions

Chapter 6 noted that database design for GLKN data will include a means for recording the protocol version under which each piece of data was collected (see Section 6.5). The metadata document will contain information about protocol versions used to collect the data.

7.8. Non-Program Data Documentation

Data which are generated and/or managed outside the I&M Program, but used in analysis with GLKN data (see Chapter 5) or distributed in any manner by the NPS, require the same level of documentation as GLKN-generated data. This includes data produced under contract with the NPS. For non-Program data, the project manager will ask the originating entity for metadata. Any contract entered into by the GLKN with data producers will stipulate that FGDC-compliant metadata in HTML, ASCII, or XML format, or in a format specified by the project manager and approved by the Network data manager. The Network data manager will assist the project manager in metadata acquisition by providing tools, format protocols, and file transfer services.

7.9. Derived Data Documentation

Several of the GLKN Vital Sign monitoring projects will produce datasets derived from data collected and certified by other entities. When a derived dataset becomes a data product, metadata are required for that dataset. Like regular project data, derived data products may not be released without proper documentation. Metadata for any such derived data will include a data lineage pointing to parent datasets.

7.10. Legacy Data Documentation

Data acquired via data mining or other means does not enjoy a level of stewardship afforded project data. When older data that have no current steward (typically a project manager) are found to be useful, metadata for those data must be established. Also, as noted in Chapter 9, no data may be released without accompanying metadata. If ancillary data are used in the creation of derived data products (for example, ancillary data are used in an analytical comparison to project data), metadata for those data are necessary to fully document the derived data product.

Credits

This chapter was adapted from concepts and material developed by Doug Wilder (Central Alaska Network).

8. DATA ANALYSIS AND REPORTING

In this chapter, we describe approaches to how data collected by the GLKN will be analyzed, summarized, and reported. Various reports and other products of the monitoring effort are described, including what they will include, who the intended audience is, how often they will be produced and in what format, and who is responsible for these products. Because the Network is actively planning for monitoring data acquisition, this chapter builds analysis and reporting approaches around anticipated monitoring data. However, general concepts described in this chapter may be applied to any GLKN project.

More in-depth discussion of the Network's analysis and reporting cycles and processes can be found in the GLKN Phase 3 Monitoring Plan (Route and Elias 2005).

8.1 Monitoring Reporting

Each Vital Sign protocol establishes requirements for data summary, analysis, and reporting. Queries and reports are integrated into the reporting function of most Vital Signs databases. Monitoring protocols also include spatial data analysis as an essential part of data reporting. The graphical presentation of locations, spatial relationships, and patterns of change over time can be powerful tools for both analyzing data and communicating results to a broad audience.

The primary role of data management staff is to prepare data as needed for analysis and reporting in coordination with the project manager.

8.1.1 Reporting Process

All data sets will go through a similar process as they move from data tables to analysis to final reports and public distribution. After data entry is completed for a season or year, data will be verified and validated. Project managers are then responsible for analyzing the data, preparing a summary of findings, and reporting on the project. In limited cases, such as with climate, the data management staff will perform some of these functions. The report and annual data set are then archived at the GLKN and posted in the appropriate locations for public access (see Chapter 9).

In addition to annual reports, periodic protocol review reports will also be produced by the Network. Within the first three years of protocol implementation, and every five years thereafter, the overall quality of the protocol will be evaluated in terms of its implementation, effectiveness, and compliance with data management standards.

Chapter 7 of the Network's *Long-term Ecological Monitoring Plan* (Route and Elias 2005) outlines in detail the Network's analysis and reporting schedule and procedures, including the associated roles and responsibilities.

As the GLKN monitoring program develops, this chapter of the Data Management Plan will serve as a timeline and guide for the data management staff on established reporting schedules and procedures. The preliminary draft reporting table (Table 8.1) will be further developed and updated to reflect additions and changes to reporting requirements and procedures.

8.2 Other Reporting

While reports on monitoring activities will constitute the majority of the reporting tasks, summaries of data that are managed in national Inventory and Monitoring (I&M) databases will also be provided to parks on a regular basis. The data compiled in NatureBib, Dataset Catalog, and NPSpecies are accessible via their respective websites; however, formatted and printed reports help ensure the information is on-site at parks, in a format accessible to all, and easy to review.

Table 8.1 Preliminary data management reporting schedule. Details of schedule, including tasks specific to each protocol, will be completed as monitoring program is implemented.

Report Type	Expected Frequency	Data Requirements
Monitoring reports		
Initial Protocol Reviews	Within 1-3 years	Review and document compliance with data entry, QA/QC, retrieval, and archiving standards. Correct data management flaws
Subsequent Protocol Reviews	Every 5 years	Review and document compliance with protocols. Refine data management procedures
Annual Reports for specific protocols	Annual	Database queries and reports per specifications from project manager. Provide number of samples and sites, relevant attributes. Document data management activities (e.g., database updates, QA/QC changes)
Analysis and Synthesis Reports	Every 3-10 years	Database queries and reports per specifications from project manager
Program review reports	Every 5 years	Evaluate effectiveness of reports, ability to share data, and utility of data products
Journal articles, book chapters, white papers	variable	Document and communicate findings and advances in data management
Symposia, workshops, conferences	variable	Present information on specific topic or subject area. Identify emerging issues and new ideas, document, and communicate findings
Other Scheduled Reports		
Metadata reports	Biennial or as needed by a park	Ensure that data mining is up-to-date at Network parks. Format and print hard copy summaries of metadata records for each park
NPSpecies report	Annual or as needed by a park	Ensure that new data are entered into NPSpecies. Format and print updated vertebrate and vascular plant species lists for each park
NatureBib report	Biennial or as needed by a park	Ensure that data mining is up-to-date at Network parks. Format and print hard copy bibliography

Credits

This chapter was adapted from concepts and material developed by Margaret Beer (Northern Colorado Plateau Network).

9. DATA DISSEMINATION

Data Management at the GLKN aims to ensure that:

- Data are easily discoverable and obtainable by all stakeholders
- Only data subjected to complete quality control are released, unless necessary in response to a Freedom of Information Act (FOIA) request
- Distributed data are accompanied by appropriate documentation
- Sensitive data are identified and protected from unauthorized access and inappropriate use

The GLKN will continue to provide links to public data products via the GLKN public websites. This will be the Network's portal for distribution of the most current data products available as the monitoring program develops. Distribution instructions for each dataset will be provided in the respective metadata.

9.1. Mechanisms for Distribution

The Network's data products (final deliverables or periodic milestones) will be 'packaged', where possible, and made available for distribution as a complete set. This 'package' is similar to the materials consolidated for natural history archiving and curation (See Chapter 10).

The GLKN plans to use the internet for exchange of dynamic and on-going data that may not be packaged as a planned product. Using the Network's IMS site, authorized users will be able to get access to and download data sets that can be created 'on demand' using a query interface.

Access to GLKN data products will be facilitated via a variety of means that allow users to browse, search, and acquire Network data and supporting documents (Table 9.1).

These means include, but are not limited to:

- Links to public data products maintained on the GLKN Network website and IMS gateway.
- NR-GIS Data Store. Distribution instructions for each dataset will be provided in the respective metadata.
- Service-wide databases, such as NPSpecies, NatureBib, and NPSTORET.
- Regional, Network, or park data servers protected with read-only access.
- External repositories, such as the Midwest Regional Climatic Center, and possibly others.
- FTP sites, CDs, or DVDs as appropriate.

Information will be made available to two primary audiences, public and NPS employees, as determined by data sensitivity and development status. Only fully documented, certified, non-sensitive data and data products may be uploaded to public distribution repositories or otherwise released to the public.

Table 9.1. Primary repositories for GLKN information and associated specimens.

Item	Repository
reports (public) • digital • hardcopy • bibliography	NPS Focus, Data Store, GLKN website and IMS Gateway Park and Network libraries, park archives NatureBib
• Network-generated digital datasets and data products (public, non-sensitive) • certified data and data products (including photos) • metadata	GLKN website, GLKN IMS Gateway, NR-GIS Data Store, NPSpecies, NPSTORET
• Network-generated digital datasets and data products (NPS staff, sensitive) • raw, validated, and analyzed data • metadata • submitted reports • digital photos • digital presentations	GLKN intranet website; selected GLKN data may be housed externally with an established MOU.
• project product materials • specimen vouchers • photograph film	Park archives or other curatorial facility (according to project protocol)
Project administrative records or miscellaneous items (hard copy)	GLKN office, NPS archive facility

The Network anticipates that an Internet Mapping Service, (IMS), which presents information in a spatial (geographic) context will be an easy and intuitive way for users to access Network monitoring information, and plans to use this tool as one method for distributing data over the internet (Figure 9.1).

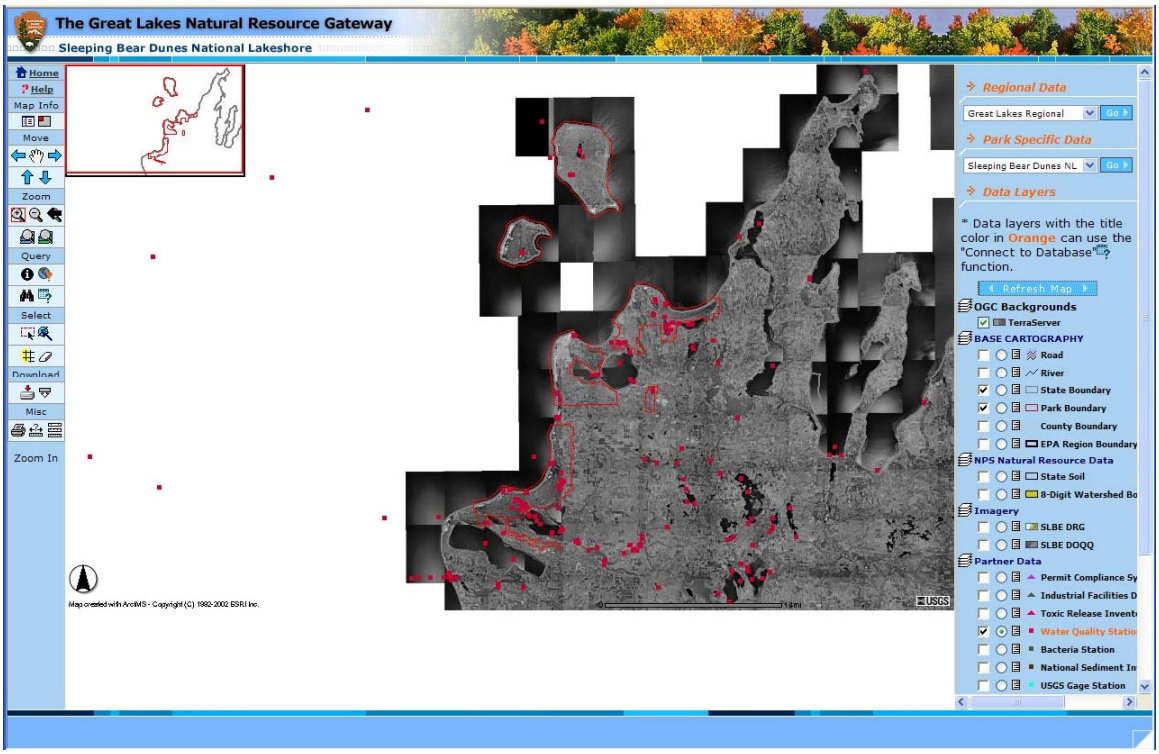


Figure 9.1. An example screenshot showing the IMS interface from a beta version of GLKN's IMS Natural Resource Data Gateway. The map shown is of water quality sample sites (red dots) at Sleeping Bear Dunes National Lakeshore, which is shown using aerial photographs.

The national water quality database (NPSTORET) maintained by the NPS Water Resources Division provides a good example of data transfer from a Network project to an on-line master database, from which data and summary products are made widely available via a web-based clearinghouse service (Figure 9.2).

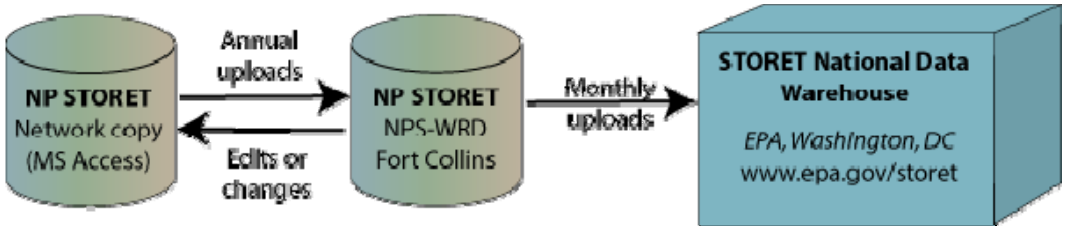


Figure 9.2. Schematic of data flow for transferring water quality data from a Network to the national master database, and for Network and public access to that master database via an on-line clearinghouse.

9.2. Ownership, FOIA, and Sensitive Data

GLKN products are considered property of the NPS (OMB, Circular A-110, Section 36).

The Freedom of Information Act, 5 U.S.C. § 552, referred to as FOIA, stipulates that the United States government, including the National Park Service, must provide access to data and information of interest to the public, regardless of whether or not the federal

government created the records. FOIA is intended to establish a right for any person to access federal agency records that are not protected from disclosure by exemption or by special law enforcement record exclusions. Under the terms of FOIA, agencies must make non-protected records available for inspection and copying in public reading rooms and/or from the internet.

The NPS is directed to protect information about the nature and location of sensitive park resources under one Executive Order and four resource confidentiality laws:

- Executive Order No. 13007: Indian Sacred Sites
- National Parks Omnibus Management Act (NPOMA; 16 U.S.C. 5937)
- National Historic Preservation Act (16 U.S.C. 470w-3)
- Federal Cave Resources Protection Act (16 U.S.C. 4304)
- Archaeological Resources Protection Act (16 U.S.C. 470hh)

When any of these regulations are applicable, public access to data can be restricted. If disclosure could result in harm to natural resources, the records may be classified as 'protected' or 'sensitive' and information withheld from the public. The following resources are recognized as sensitive by the NPS:

- endangered, threatened, rare, or commercially valuable NPS resources
- mineral or paleontological sites
- objects of cultural patrimony
- significant caves

The Network will comply with all FOIA restrictions regarding the release of data and information as instructed in NPS Director's Order #66 and accompanying Reference Manuals 66A and 66B (currently in development). Managing natural resource information that is sensitive or protected requires the following steps:

- Identification of potentially sensitive resources
- Compilation of all records relating to those resources
- Determination of which data must not be released in a public forum
- Management and archival of those records to avoid their unintentional release

Classification of sensitive data will be the responsibility of Network staff, park superintendents, and project managers. Network staff will classify sensitive data on a case-by-case, project-by-project basis and will work closely with project managers to ensure that potentially sensitive park resources are identified, that information about these resources is tracked throughout the project, and that potentially sensitive information is removed from documents and products that will be released outside the Network.

Consider the following questions to determine whether information should be protected:

- Has harm, theft, or destruction occurred to a similar resource on federal, state, or private lands?
- Has harm, theft, or destruction occurred to other types of resources of similar commercial value, cultural importance, rarity, or threatened or endangered status on federal, state, or private lands?

- Is information about locations of the resource in the park specific enough so that the park resource is likely to be found at these locations at predictable times now or in the future?
- Would information about the nature of the park resource, which is otherwise not of concern, permit determining locations of the resource if the information were available in conjunction with other specific types or classes of information?
- Even if relatively out-dated, is there information that would reveal locations or characteristics of the park resource such that the information could be used to find the park resource as it exists now or is likely to exist in the future?
- Does NPS have the capacity to protect the park resource if the public knows its specific location?

9.2.1. Access Restrictions on Sensitive Data

Sensitive park resources will be identified through a cooperative effort. The Network and park staffs should identify all potentially sensitive park resources to the manager for each project. Reciprocally, the project manager must identify any known references to potentially sensitive park resources.

Network staff are responsible for managing access to sensitive data handled by the Program. All potentially sensitive park resources will be identified and investigators working on Network projects will be informed that:

- All data and associated information must be made available for review by Network staff prior to release in any format; and
- Any information classified as protected should not be released in any format except as approved in advance by the National Park Service.

When preparing information for any repository, the Network staff ensures that all protected information is properly identified and marked. All references to protected information are removed or obscured in any reports, publications, maps, or other public forum.

Network staff will remove any sensitive information from public versions of documents or other media. They will isolate sensitive from non-sensitive data and determine the appropriate measures for withholding sensitive data. The main distribution applications and repositories developed by the I&M Program are maintained on both secure and public servers and all records marked ‘sensitive’ during uploading will only be available on the secure servers. Procedures for assigning a sensitivity level to specific records when uploading to both the NPSpecies and NatureBib databases are given on the following websites:

- *http://science.nature.nps.gov/im/apps/npspp/index.htm*
- *http://www.nature.nps.gov/nrbib/index.htm*

Thus, access to data on sensitive park resources can be limited to Network staff or research partners. However, limits to how these data are subsequently released must also be clearly defined. It is crucial that the person(s) uploading records to the online applications (repositories) is familiar with the procedures for identifying and entering protected information.

9.2.2. NPS Only versus Public

Only data subjected to complete quality control are released, unless release is necessary in response to a FOIA request. Products of incomplete, poor, or questionable quality (typically legacy data) may not be appropriate for the public and should be managed separately from those of acceptable quality. These lower quality products may be the only source of information on the natural history of the park, may have been the basis for early management decisions, and may still be valuable in-house.

Provided low quality products do not contain sensitive data (as described earlier in this chapter), these data may be released to the public upon specific FOIA request. They must be accompanied with qualifying documentation.

National I&M Program applications, such as NatureBib and NR-GIS metadata, provide a means to flag data as "NPS Only" or "Public". Additional notes describing the quality should be added to the records in these applications, as resources permit.

9.2.3 Establishing Data Ownership: Cooperative or Interagency Agreements

To ensure that proper ownership, format, and development of Network products is maintained, all cooperative or interagency work must be conducted as part of a signed collaborative agreement. Every cooperative or interagency agreement or contract involving the GLKN must cite OMB Circular A-110 under the *Reports and Deliverables* Section of all agreements and contracts. The following shows appropriate language to use when citing Circular A-110:

> "As the performing organization of this agreement, *institution or organization name* shall follow the procedures and policies set forth in OMB Circular A-110."

Cooperative or interagency agreements or contracts must include a clearly defined list of deliverables and products. Details on formatting and media types that will be required for final submission must be included. Typical products include, but are not limited to, field notebooks, photographs (hardcopy and digital), specimens, raw data, and reports.

The following statement must be included in the Reports and Deliverables section of all GLKN agreements and contracts:

> "All reports and deliverables must follow the most recent versions of the Great Lakes I&M Network product specifications."

Cooperative agreements and contracts should also provide a schedule of deliverables that includes sufficient time for NPS review of draft deliverables before scheduled final submissions.

9.3. Data Release Policy

The GLKN will maintain the following policy regarding the release of data and related products:

1) Only fully documented, certified, non-sensitive data and data products may be uploaded to public distribution repositories or otherwise released to the public or other non-NPS recipient.

2) Any released data or data products must be accompanied by full metadata and any additional supporting documentation as determined by the project manager.
3) Data from the GLKN monitoring program will be made available according to GLKN reporting and data integration schedules specified in the most recent versions of this document (see Chapter 8) and the GLKN Monitoring Plan, (Route and Elias 2005).
4) Non-monitoring data will be released by GLKN upon completion of data certification (verification and validation) and on condition of project manager approval. However, data for which analysis and reporting have not been completed but which are otherwise certified will be released no later than one year after certification.
5) Distribution instructions for each dataset will be included in respective metadata.
6) Distribution logs specifying recipient name and contact information, intended use of data, export file format, delivery date and method, data content description, and range (by date and geography) of data delivered and description of distributed items will be kept for each distinct dataset or product managed by the GLKN.
7) At the discretion of the project manager, any reports utilizing the data may also accompany distributed data.

9.4. Feedback Mechanisms

Comments and questions concerning I&M project data are welcome at any time and may be submitted via written correspondence, e-mail or telephone to the primary contact of a project or to the network coordinator. The GLKN websites will also provide links that NPS staff, cooperators, and the public can use to provide feedback on data and information distributed as part of GLKN operations. A 'comments and questions' link will be provided on the main page of the site for general questions and comments about the Network's program and projects. A more specific 'data error feedback' link will direct comments to GLKN staff pertaining to errors found in website-accessible data. Data feedback and response activity will be made available to the Technical Committee on a regular basis.

Data Error Feedback Response Procedures

The GLKN will use the following procedure to respond to data feedback:

- Network staff (or automated process) will immediately acknowledge receipt of any feedback to the sender.
- Data error reports will be recorded in an error log associated with each database.
- Data errors will be investigated promptly following an error report submission. If appropriate and feasible, confirmed errors will be corrected immediately in all data repositories. If immediate correction is not possible, the data in question will be removed from all dissemination mechanisms.
- A report detailing the reported error and response will be prepared by the Network data manager and submitted to the project manager, Network coordinator and the error reporter. If a data error is not immediately correctable, the error response report will include recommendations for correcting the error.

- An appropriate level of communication to stakeholders will be maintained during all data error investigations.

Credits

This chapter was adapted from concepts and material developed by Gordon Dicus (Pacific Island Network). Sara Stevens (Northeast Coastal and Barrier Network), Wendy Schumacher (NPS Washington Office), Margaret Beer (Northern Colorado Plateau Network) and Doug Wilder (Central Alaska Network) also made significant contributions to the original template version material used in this chapter.

10. RECORDS MANAGEMENT AND NATURAL HISTORY ARCHIVING

This chapter applies to documents such as final reports prepared by NPS staff or contractors, program administrative documents, contracts and agreements, memoranda of understanding, and other documents related to GLKN activities. This chapter also applies to physical items such as natural history specimens, photographs, and audio tapes. In most instances these documents and objects are essential companions to the digital data described throughout this plan.

This chapter is not intended to provide a full description of archiving procedures. That information is covered in museum manuals and regulations. It is intended to provide guidance for transferring completed project deliverables to a secure and long-term storage facility managed by park and/or regional repositories.

10.1. NPS Standards

Direction for managing physical and digital materials is provided in NPS Director's Order 19: Records Management (2001) and its appendix, NPS Records Disposition Schedule (NPS-19 Appendix B, revised 5- 2003) (*http://www.nps.gov/policy/DOrders/DOrder19.html*). Director's Order NPS-19 states that all records of natural and cultural resources and their management are considered mission-critical records, that is, necessary for fulfillment of the NPS mission. NPS-19 further states:

> *Mission critical records are permanent records that will eventually become archival records. They should receive the highest priority in records management activities and resources and should receive archival care as soon as practical in the life of the record.*

Section N of NPS-19 Appendix B, which provides guidelines on records related to natural resources (including, specifically, products of Inventory and Monitoring Programs), indicates that all natural resource records are considered "permanent" and need to be retained either in an appropriate park museum facility or at the National Archives. It also indicates that non-archival copies of natural resource-related materials are "...potentially important for the ongoing management of NPS resources" and should not, in any instance, be destroyed. The NPS Museum Handbook provides the overarching guidance for archival procedures. In particular, Part II of Appendix A: Mandates and Standards for NPS Museum Collections lists the cultural and natural history laws, regulations and conventions for NPS museum collections, and should be reviewed prior to object collections.

10.2. Park Unit and Network Standards

Direction for managing materials for each park unit is described in each collection facility's "Scope of Collection Statement". In general, only materials directly "... related to one or more of the park's themes or site-related materials that the NPS is legally mandated to preserve" will be accepted. All of these materials are required to be cataloged in the NPS cataloging system (ANCS+, see 36 CFR, Section 2.5).

The GLKN will adhere to the standards described above. All specimens collected as part of GLKN projects are considered property of the respective park. Procedures for the collection and processing of specimens or samples will be dictated by curatorial staff at the park or regional level, as appropriate. In some instances, the parks may delegate the GLKN as curator for specimens that are central to the Network's research.

10.3. Role of Curators in Curation and Museum Collection Storage

Curators and archival specialists for the region and GLKN parks are an excellent source of expertise, advice, and guidance on curatorial issues, and they have a role in almost every project undertaken by the Network. Project leaders should consult with curatorial staff early in the project planning stage to ensure that all aspects of museum curation of documents, specimens, and other objects are considered, and that any associated expenses are included in project budgets.

When there are field collections associated with a park research permit, an accession number is assigned by the park curator when the permit is issued. After collection, a range of catalog numbers can be assigned by the curator for the principal investigator to use. The accession and catalog number assignments must be requested by the project leader.

The project leaders will package all project materials and give the package to the park or its designated curator, as established in their collection permit. The curator will ensure that the materials are properly stored, cataloged, and archived. The project leader is ultimately responsible for verifying that cataloging and archiving has been completed.

10.4. Workflow

Project managers are responsible for preparing materials for the curator. The workflow illustrated here and described in this chapter refers to the general procedure for making the transition between the project leaders and the museum staff (Figure 10.1). The museum staff will follow a much more detailed procedure for actual curation. The workflow schematic begins with the delivery of the products, walks through the electronic storage of these products, and then highlights the storage and archiving of materials (Figure 10.1).

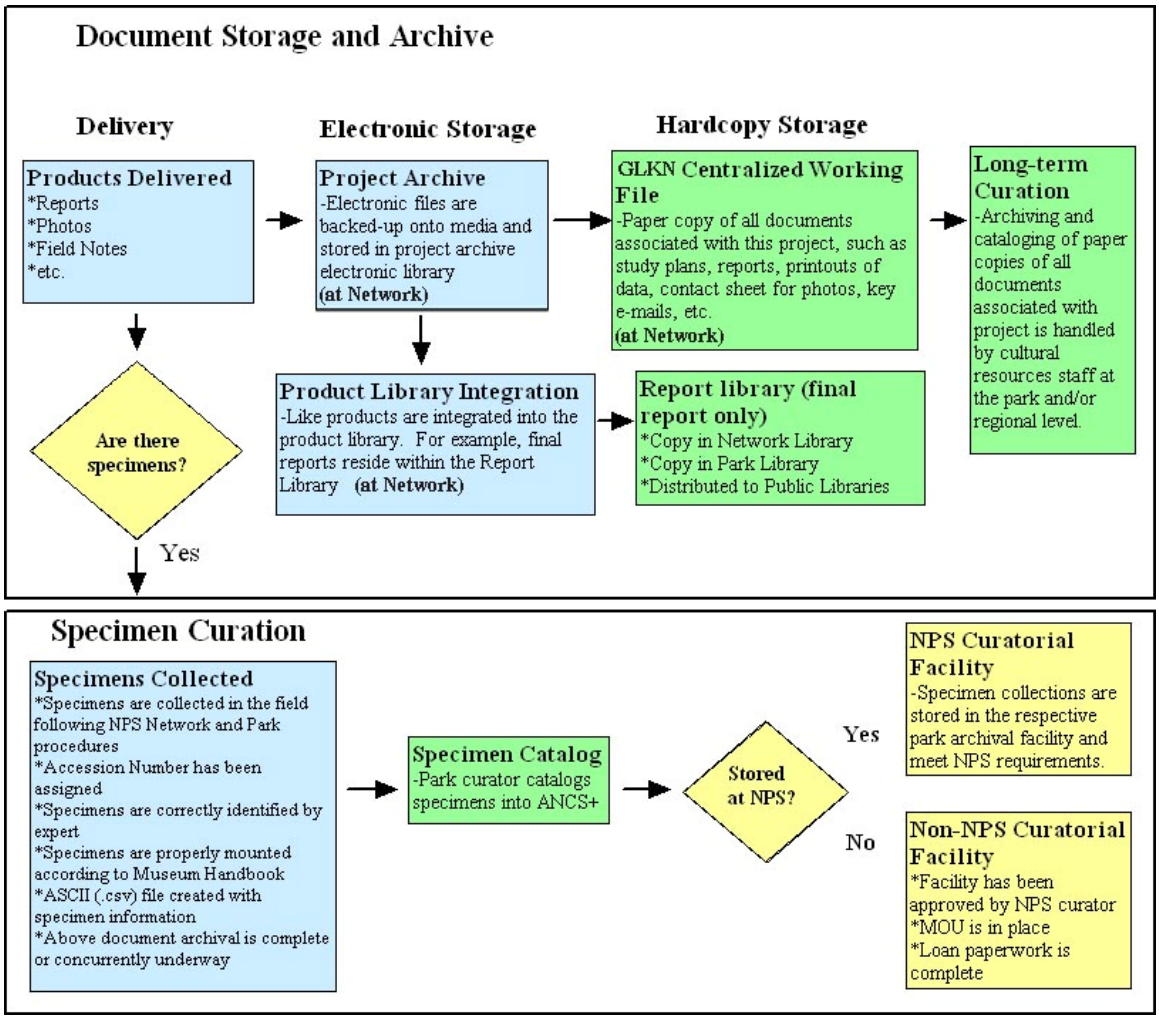


Figure 10.1. Schematic of natural history curation workflow for both hard copy documents and specimens.

10.5. Network Archiving Process

Project materials should be consolidated and packaged for curation when a project is complete, or for long-term projects, reaches a milestone. The project leader should acquire an accession number from the respective curator.

The project leader should package the project information as follows:

- Include a cover sheet, memo, or Collection Catalog Worksheet listing contact information, project abstract and purpose, sensitivity and use of materials, and all materials included in the package.
- All materials should be clearly labeled with the:
  - Park acronym
  - Date or range of dates
  - Accession number
  - Project number
- Field notes should be on acid free paper (preferred) or copy paper and stored in a 3-ring binder or book box.

- Other paper materials such as reports and data printouts should be stored on acid free paper (preferred) or copy paper and placed individually in an acid free, labeled folder.
- Photographs should be printed and documented.
- Voucher specimens should be labeled according to the Museum Handbook.
- For specimens not residing in park collection facilities, loan paperwork should be completed and a copy of the form should be stored with the project package.
- Specimens not residing in NPS repositories should be stored in a selected facility that meets NPS museum collection standards.
- Curation of specimens must be addressed in GLKN Vital Signs protocols and should specify:
  - The name and address of the storage facility
  - Memorandum of Understanding, if applicable
  - What is being curated or loaned (list of items)
- All electronic materials should be included on CD or DVD media. Jewel case should be labeled with the:
  - Park acronym
  - Date created
  - Range of dates for information
  - Accession Number
  - Project Number
  - List of contents

The museum staff will catalog this information into ANCS+ and will properly store the materials according to the Museum Handbook. If specified in arrangements with the Network and each affected park, similar packages should be prepared as "working copy" for the Network and the park(s).

10.6. Archive and Storage Locations

Many of the GLKN parks do not have adequate facilities to properly store and archive documents, photographs, artifacts, and specimens from all natural and cultural resource projects. Most GLKN parks actively store documents, but many look to other institutions for storage and archiving of artifacts and specimens. Storage of artifacts and specimens in non-NPS facilities present challenges for park staff wishing to access materials. In contrast, an institution that has ample resources committed to an organized and well-maintained collections department serves scientists and researchers well because it provides a centralized location for materials from many collection sites.

The GLKN parks will continue to cooperate with various regional museums or curatorial facilities to archive artifacts and specimens collected in the parks.

10.6.1. Network Working Files and Library

While a GLKN project is on-going, all relevant materials (reports, administrative documents, critical e-mails or correspondence, and data print outs) will be stored in the GLKN offices in an organized manner. Upon project completion, the GLKN will submit all final and pertinent project materials to the appropriate park for long-term storage and

archiving. If the park lacks the resources to store and archive these materials, then the GLKN will provide long term storage. In the future, the GLKN will establish a relationship with an NPS or NPS-sanctioned archival facility within the Midwest Region to store the mandated documents.

A separate section of the GLKN library will contain annual and final project reports and may contain additional reports that are related to inventories or Vital Signs monitoring. All materials in this library will be cataloged into the NatureBib database and should be citable. This section of the GLKN library should not contain draft materials.

10.7. NPS Automated National Catalog System (ANCS+)

The NPS Service-wide cataloging system is ANCS+ (using ReDiscovery software customized for NPS). Park staff with cataloging duties are typically the only individuals with access to this software. Park collections are entered into ANCS+ at the local level, then "rolled up" to the Service-wide database.

The cataloging system uses an accession number to identify a collection and catalog numbers to identify the items within the collection. If specimens were collected, the project will have a single accession number, but will contain multiple catalog numbers (one for each specimen collected). Park staff may access the catalog system by one of three methods:

- Contact a park curator and review together
- Request a copy of the program and data from the park curator
- Access the information online using ReDiscovery Web Catalog (http://www.nps.gov/curatorial)

Uploading of a park's ANCS+ records to the online ReDiscovery Web Catalog is provided at additional cost to the park. It should be noted that the online catalog may not represent all collection holdings at a given park.

10.8. Photographs

Museum curators have been reluctant to fully embrace digital photography and some have expressed concern that, with the accelerating rate of technological change, documentary heritage is in danger of being lost in the information age (Cox 2000).

Photographic methods for any given project will be developed by the project manager and GLKN staff. The GLKN accepts both digital and film-based photographs. Under the project manager's direction, the GLKN will screen photographs for quality, redundancy, and adequate documentation. For projects using analog methods, the Network will accept 4x6 prints or 35mm slides (preferably Kodachrome or Ektachrome). The long-term stability of 35mm slides has been documented by Wilhelm and Brower (1993). Original photographs are high priority items for archival and long-term storage.

Slides should be labeled using indelible pigment ink or using laser-printed archival-quality slide labels. Slide labels should include: a unique ID, project name, accession number, photographer name, photograph date, a brief identification of contents (e.g., species name, plot ID), and geographic location (i.e., coordinates and datum or a description). All slides should be stored in polypropylene slide sleeves. In addition, all

slides will be scanned and saved digitally as Tagged Image File Format (TIFF) (preferred) or Joint Photographic Experts Group (JPEG) files, which will be used as the primary means of distributing or reproducing the images.

Photographic prints should be stored in individual polypropylene sleeves and within archival boxes. Each photograph should be labeled on the back, using an archival-quality label. Labels should include the same information elements required for slides, and may be either laser- or hand-printed. If a contractor is submitting photographs, corresponding TIFF files must also be submitted.

Every image, regardless of format, will have an entry into a photographs database (ThumbsPlus or other GLKN-approved database) where attributes such as the electronic file name, keywords, project, photo description, photographer, date, and location are cataloged.

10.9. Specimens

Planning for monitoring projects includes budgeting for specimen preparation, records processing, cataloging, and storage. Specimens collected under the auspices of the GLKN will be provided to the respective park in which they were collected for curation, or to a repository approved by a park. Project managers are responsible for providing park curators with necessary data for cataloging each specimen. Network data management staff will assist project leaders and curators to ensure that specimens records are in comma-delimited format (.csv) for automated uploading into ANCS+. Data provided to non-NPS curators will be in an appropriate format specified by the institution.

10.10. Other Materials

All materials related to a project should be included in the collection. For materials (e.g., audio and visual) not described in this chapter, project managers should discuss archival requirements with the appropriate curatorial staff. The backup and archiving of digital material is cover earlier in this document (see Chapter 4).

Credits

This chapter was adapted from concepts and material developed by Gordon Dicus (Pacific Island Network). Margaret Beer (Northern Colorado Plateau Network) and Rob Daley (Central Alaska Network) also made significant contributions to the original template version material used in this chapter.

11. IMPLEMENTATION

As a data provider, our data management processes may be satisfactory and may meet our needs for a number of years. The I&M Program, however, focuses attention on our data consumers – managers who will make decisions based on the data we provide or researchers who will use these data decades from now. The central goal of Vital Signs monitoring is to understand the long-term patterns and processes of park ecosystems at multiple spatial scales. Good data management practices are essential to achieving this goal. It costs less to save data than to collect it again, and ecological data often cannot be collected again at any cost.

Some of the practices in this Data Management Plan may be new to staff and principal investigators. However, with few exceptions, the DMP does not include any new requirements. Almost every requirement comes from law, Director's Orders, or the I&M Program. The DMP helps to put these requirements into context and provides operational guidance to fulfill these requirements. The DMPs for each of the 32 I&M networks are the first comprehensive documents of their kind and will establish the means by which managers, educators, and the greater scientific community can reach data and information about network and park resources.

Good data management practices take time. Some Vital Sign collection procedures and data management practices are already in use and may require minimal revisions. Other data management practices may involve several iterations before the procedures and databases are acceptable and functional.

11.1. Common Data Management Sins

Data mining provides a means of discovering existing data and is an important component of all projects (Chapter 5). The Network has put considerable effort into mining data and information useful to the I&M Program. This consisted of looking through archives, office filing cabinets, natural resource folders, libraries, and electronic files. Not surprisingly, some data and information was well-maintained and documented, and some was not. Examples of common data management problems are:

- *Unfinished projects*. Failure to finish a project can happen for a number of reasons beyond the staff's control. Unfinished projects, however, are the most difficult to manage. It takes longer to document an incomplete project than a finished one. The best data management practice is to finish the project or reach a logical milestone.
- *Lack of header information*. Often, sheets of information do not have a date, name or subject. Reading the material and determining if the information is relevant is extremely time consuming. Relevant information is often discarded because it does not have adequate header information. A good data management practice is to use a memorandum format or a standard header for all files for a project, no matter how insignificant.
- *Lack of documentation*. Most projects and data are inadequately documented. Without documentation, the data are often considered unusable. Documentation

techniques, which should be on-going throughout the duration of the project are described in this document.

- *Lack of file naming standard and version control*. In many cases it was impossible to know which electronic file contained the final product.
- *No archive or final storage*. Data tucked away in an individual's desk or personal computer isn't accessible to anyone else. It often requires in-house knowledge to know that the information even exists. When there is a turnover in staff, this information is often lost.

11.2. Years 1-5 of Implementation

The GLKN is just beginning to implement this data management plan. The first few years of implementation will involve learning, testing, and refining. During the coming years, the DMP will be implemented for the first Vital Signs monitoring protocols and data management training will be provided to Network and park staff. Templates and SOPs will be developed and tested, and will then be applied to the other Vital Signs monitoring protocols. Some bottlenecks are anticipated; identifying and eliminating problems will be an important part of implementation.

Implementation Goals for Years 1-5:

The following outcomes should be achieved within the first five years of this Plan's implementation:

- All Network staff understand the fundamentals of data management, which include:
  - File management,
  - Documentation,
  - Quality assurance and quality control,
  - Electronic storage, and
  - Archive storage.
- Data management practices will be improved by implementing:
  - Conceptual Data Models for Vital Signs monitoring protocols, and
  - Testing of data entry prior to field work.
- Common SOPs that can be used for multiple protocols will be developed
- Data management staff will inform and direct the data management aspects of the first Vital Signs monitoring protocols to be implemented

11.3. Years 5+ of Implementation

In five years, the DMP will be revised. This next version may be more streamlined and direct. With prior experience and commonplace practices, generalizations may be eliminated.

Implementation Goals for Years 5+:

- The DMP will be revised and streamlined for Vital Signs monitoring.
- All implemented Vitals Signs monitoring protocols will following the DMP guidance
- Databases and custom reporting will continue to be enhanced.

- Framework and gateway for integration of data with other agencies or networks will be established.

Credits

This chapter was adapted from concepts and material developed by Gordon Dicus (Pacific Island Network).

LITERATURE CITED

Cox, R.J. 2000. Searching for authority: archivists and electronic records in the New World at the fin-de-siècle. First Monday 5:1. (http://firstmonday.org/issues/issue5_1/cox/index.html).

Edwards, D. 2000. Data quality assurance. Pages 70-91 *in* Michener, W.K., and J.W. Brunt, editors, 2000, Ecological data: design, management and processing. Blackwell Science Ltd. Methods in Ecology Series, Malden, MA.
.
Palmer, C.J., and E.B. Landis. 2002 Draft. Lake Mead National Recreational Area Resource Management Division: Quality System Management Plan for Environmental Data Collection Projects.

Route, B. and J. Elias (editors). 2005. Phase 3 Report: Long-term Ecological Monitoring Plan for the Great Lakes Network. U.S. National Park Service, Great Lakes Inventory and Monitoring Network. Technical Report GLKN/05/19.

Southwould, A. 2002. Recommended Database Strategies, including I&M Database TemplateVersion 1.0. Draft. Unpublished manuscript.

Wilhelm, H., and C. Brower. 1993. The permanence and care of color photographs: traditional and digital color prints, color negatives, slides, and motion pictures. Preservation Publishing Company, Grinnel, IA.

ACKNOWLEDGEMENTS

The creation of this plan involved many individuals both within and without the National Park Service. The Data Managers from the first group of twelve I&M networks to complete their planning process developed a template and drafts of the chapters that have been included in this Plan. Their contribution was invaluable. Specific credits for this work are included at the end of each chapter. Undoubtedly some individuals that made substantial contributions to these initial chapter drafts were overlooked in this documents chapter credits, we apologize for these omissions and extend out gratitude to all

A special acknowledgement is due to John Boetsch for coordinating the collective writing effort amongst the “first tier” network data managers, who contributed greatly to the content and completion of this plan.

Great credit and thanks is due to Gordon Dicus (Pacific Island Network) for making a draft of the PACN data management plan available to other data managers, and to Dorothy Mortenson for her efforts to establish cooperation amongst the “second tier” data management plan drafters.

The following GLKN staff that provided key review and critique: Bill Route, Tammy Keniry, Joan Elias, and Jennifer Sieracki.

APPENDIX A: BASIC RESOURCE INVENTORIES AND REPOSITORIES FOR THE GLKN.

Inventory	Information Access
Natural resource bibliography	NatureBib (www.nature.nps.gov/nrbib)
Base cartographic data	http://science.nature.nps.gov/nrdata/
Geology map	NPS Geologic Resources (www2.nature.nps.gov/geology/inventory/gre_pub2.htm)
Soils map	http://websoilsurvey.nrcs.usda.gov/app/WebSoilSurvey.aspx
Meteorological data	Midwestern Regional Climate Center (http://mrcc.sws.uiuc.edu)
Air quality	NPS Air Resources (www2.nature.nps.gov/air/)
Air quality related values	NPS Air Resources (www2.nature.nps.gov/air/)
Water body location and classification	NPS Water Resources (www.nature.nps.gov/water)
Water quality data	NPS Water Resources (www.nature.nps.gov/water)
Vegetation map	GLKN (www.nature.nps.gov/im/units/GLKN/PlantInventory.htm)
Documented species occurrence of vertebrates and vascular plants	NPSpecies (science1.nature.nps.gov/npspecies) and GLKN (www.nature.nps.gov/im/units/GLKN/Inventory.htm)
Species distribution and status of vertebrates and vascular plants	NPSpecies (science1.nature.nps.gov/npspecies) and GLKN (www.nature.nps.gov/im/units/GLKN/Inventory.htm)

APPENDIX B: GLKN DATA MANAGEMENT PLAN REVISIONS

The latest version of the DMP is available on the GLKN website (*www.nature.nps.gov/im/units/GLKN/DataMgt.htm*) and will include a revision log as an appendix.

Revision and maintenance of the DMP will be incremental as needed to reflect best practices and current information. The data management work group (See Section III.2) for the GLKN will be responsible for this level of on-going maintenance. Recommendations for changes may be forwarded to any member of the data management work group by any interested party or user of GLKN Inventory and Monitoring data. These recommendations will be discussed at regular meetings of the data management work group. Changes to the DMP will be made at the discretion of the data management workgroup only after thorough review and assessment of anticipated ramifications. The amount of time required for adequate review and assessment of ramifications will likely vary greatly in accordance with the nature of the suggested change. When enough revisions have been made to merit an additional round of review, this process will be taken up with the network technical committee. Otherwise, the plan will be scheduled for a more formal revision and review on a regular basis (at least every 5 years).

The titles of updated or revised sections of the plan will be followed by brackets containing the date of the latest version of that section and a brief update history. For example, the title of Section II.2 may be followed by "[06/02/2006; Update History: 1/15/2005, 10/23/2004]". This indicates that the current wording of that section was last updated on June 2, 2006 and that it was previously revised on January 15, 2005 and October 23, 2004. The list of previous dates will be limited to two dates followed by a "…" if there are additional dates. A "Document History" will be maintained as part of Appendix A (this appendix) of the Data Management Plan which will provide further details.

Document History:

12/15/05 Mark Hart

The first draft of the body of this document was completed in draft form and submitted to the WASO I&M Program for review. Mark Hart took the lead role in formulating the initial plan, with substantial assistance from Ulf Gavfert.

APPENDIX C: NATIONAL-LEVEL INVENTORY AND MONITORING INFORMATION MANAGEMENT STRATEGIES

The need for effective natural resource information management cuts across NPS divisional boundaries and management strategies must be defined at the highest level possible. In this context, integrated inventory and monitoring of natural resources is multidisciplinary and requires national-level, programmatic data and information management strategies for success.

The basic strategy of natural resource and therefore inventory and monitoring information management is to provide integrated natural resource databases and information systems that enhance NPS managers' and staff's access and use of timely and valid data and information for management decisions, resource protection, and interpretation. Inventory and monitoring information needs are broadly separated into two categories:

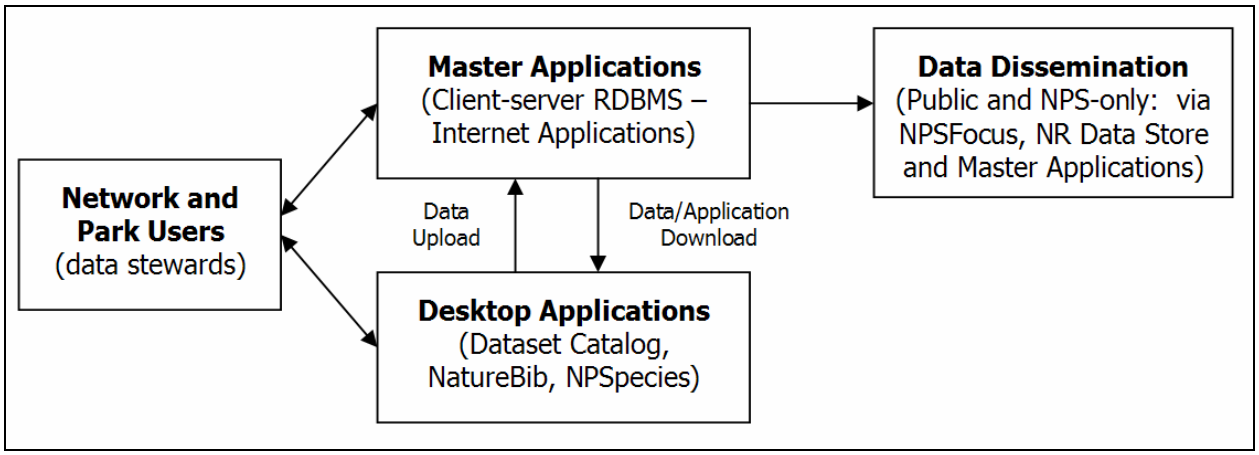
- *Detailed data and information needed for onsite resource management and protection*. The information used to guide natural resource management decisions must be specific to inform and be useful to management staff at parks and central offices.
- *Summary information needed to describe the resources and their condition*. This kind of information usually needs to be aggregated across the National Park Service for use by NPS and DOI managers and central office personnel to answer requests from Congress and for budget, program, and project planning.

The NPS Natural Resource Program Center (NRPC) and the I&M Program actively develop and implement a national-level, program-wide information management framework. NRPC and I&M staff integrate desktop database applications with internet-based databases to serve both local and national-level data and information requirements. NRPC staff members work with regional and support office staff to develop extensible desktop GIS systems that integrate closely with the database systems. Centralized data archiving and distribution capabilities at the NRPC provide for long term data security and storage. NRPC sponsors training courses on data management, I&M techniques, and remote sensing to assist I&M data managers with developing and effectively utilizing natural resource information.

National-level application architecture

To achieve an integrated information management system, three of the national-level data management applications (NatureBib, NPSpecies, and NR-GIS Metadata Database) utilize a distributed application architecture with both desktop and internet-accessible (master) components.

Model of the national-level application architecture.



NatureBib

NatureBib is the master database for bibliographic references that merges a number of previously separate databases such as Whitetail Deer Management Bibliography (DeerBib), Geologic Resource Bibliography (GRBib), and others. It also contains citation data from independent databases like NPSpecies and the Dataset Catalog and NR-GIS Metadata Database. It currently focuses on natural resource references, but may eventually be linked to references on cultural resources and other park operations. As with NPSpecies and NR-GIS Metadata Database, it is possible to download data from the master web version into the MS Access desktop version that can be used locally on computers with limited internet connectivity *(http://www.nature.nps.gov/nrbib)*.

NPSpecies

NPSpecies is the master species database for the NPS. The database lists the species that occur in or near each park, and the physical or written evidence for the occurrence of the species (e.g., references, vouchers, and observations). Taxonomy and nomenclature are based on ITIS, the interagency Integrated Taxonomic Information System. The master version of NPSpecies for each park or network can be downloaded from the master website into an MS Access version of NPSpecies. The internet-based version is the master database, which can be accessed via password-protected logins administered by park, network and regional data stewards assigned for each park and network. The master database requires that species lists are certified by networks before any data will be available to the public. NPSpecies is linked to NatureBib for bibliographic references that provide written evidence of a species' occurrence in a park and will be linked to NR-GIS Metadata Database to document biological inventory products. The MS Access application and additional details can be found at the NPSpecies website *(http://science.nature.nps.gov/im/apps/npspp/index.htm)*.

Biodiversity Data Store

A digital repository of documents, GIS and other data sets that contribute to the knowledge of biodiversity in National Park units, including presence/absence,

distribution and abundance (*http://science.nature.nps.gov/im/inventory/biology/index.htm*).

NR-GIS Data Store

As with other Service-wide applications, the master metadata database (NR-GIS Metadata Database) is available through a website and is linked to NPSpecies (the NPS species database) and NatureBib (the bibliographic database). It will be possible to download a version in MS Access format from the master website (*Dataset Catalog: http://science.nature.nps.gov/im/apps/datacat/index.htm and NR-GIS Metadata Database: http://science.nature.nps.gov/nrdata).*

Dataset Catalog is a desktop metadata database application developed by the I&M Program to provide a tool that parks, networks, and cooperators can use to inventory and manage data set holdings. Although not designed as a comprehensive metadata tool, the Dataset Catalog is used for cataloging abbreviated metadata about a variety of digital and non-digital natural resource data sets. The Dataset Catalog helps parks and networks begin to meet Executive Order 12906 mandating federal agencies to document all data collected after January 1995. It provides brief metadata and a comprehensive list about all resource data sets for use in data management, project planning, and more stringent metadata activities.

Other National-Level Inventory and Monitoring Information Management and GIS Applications

NPSTORET

STORET is an interagency water quality database developed and supported by the Environmental Protection Agency (EPA) to house local, state, and federal water quality data collected in support of managing the nation's water resources under the Clean Water Act. STORET is used by NPS as a repository of physical, chemical, biological, and other monitoring data collected in and around national park units by park staff, contractors, and cooperators. The NPS operates its own Service-wide copy of STORET and makes periodic uploads to the EPA STORET National Data Warehouse so that data collected by and for parks will be accessible to the public. NPS Director's Order 77 indicates that the NPS should archive water quality data in STORET, and the NPS Water Resources Division (WRD) requires that any data collected as part of a funded WRD project get archived in STORET. NPSTORET (also known as Water Quality Database Templates) the NPS master database designed to facilitate park-level standardized reporting for STORET. The database is still in development, but metadata, protocols, data dictionaries, and reporting capabilities are available through a front-end form. Upon implementation, network staff and cooperators will be able to use the MS Access version of NPSTORET either as a direct database for data entry and management, or as a means of submitting data for upload to STORET by WRD staff. The MS Access application and additional details can be found at: *http://www.nature.nps.gov/water/infodata.htm.* Additional information on STORET can be found at: *http://www.epa.gov/storet.*

Natural Resource Database Template

The Natural Resource Database Template (NRDT) is a flexible, relational database in MS Access for storing inventory and monitoring data (including raw data collected during field studies). This relational database can be used as a standalone database or in conjunction with the GIS software (e.g., ArcView or ArcGIS) to enter, store, retrieve, and otherwise manage natural resource information. The template has a core database structure that can be modified and extended by different parks and networks depending on the components of their inventory and monitoring program and the specific sampling protocols they use. The Natural Resource Database Template is a key component of the I&M program's standardized monitoring protocols. These monitoring protocols include separate modules detailing different aspects of monitoring project implementation, from sampling design to data analysis and reporting, and include data management components that describe database table structure, data entry forms and quality checking routines. Approved monitoring protocols, including the databases that are based on the Database Template, are made available through a web-based protocol clearinghouse (see below). A description of the Database Template application, a data dictionary, and example implementations are located on the NR Database Template website *(http://science.nature.nps.gov/im/apps/template/index.htm).*

Natural Resource Monitoring Protocols Clearinghouse

The Natural Resource Monitoring Protocol Clearinghouse (i.e., Protocol Database) is a web-based clearinghouse of sampling protocols used in national parks to monitor the condition of selected natural resources. The database provides a summary of, and in many cases allows the user to download a digital copy of, sampling protocols that have been developed by the prototype monitoring parks or other well-established protocols used in national parks. The Protocol Database also makes it possible to download database components (e.g., tables, queries, data entry forms) in MS Access that are consistent with the Natural Resource Database Template that have been developed for a particular protocol. See the Protocol Database website for available protocols (*http://science.nature.nps.gov/im/monitor/protocoldb.cfm*).

NR-GIS Data Store

The NR-GIS Data Store is a key component of the data dissemination strategy employed by the I&M Program. The NR-GIS Data Store is a graphical search interface that links dataset metadata to a searchable data server on which datasets are organized by NPS units, offices and programs. The interface allows customized public or protected searches of natural resource datasets, inventory products and GIS data produced by the I&M and Natural Resource GIS Programs. Each park or network is able to post and curate its data on the server. The NR-GIS Data Store will be integrated with the master NR-GIS Metadata Database application to streamline programmatic data documentation and dissemination processes. The simple browse function of this server can be accessed at: *http://nrdata.nps.gov/*.

See the NR-GIS Data Store website for further information (*http://science.nature.nps.gov/nrdata*).

APPENDIX D: PROJECT INFORMATION MANAGEMENT PROCESS AND WORK FLOW: DELIVERABLES, REPOSITORIES, AND GUIDANCE DOCUMENT REFERENCES

Project phase	Deliverables	Repository	Data management guidance
Planning and approval	• proposal (optional) • request for bids (optional) • study plan, work plan • permit application and permit • contract or agreement • record in accounting db	• park project funding requests in PMIS/RAMS • permit applications in RPRS • network project management database • working project files, then document archives • final copies of study plan sent to document archives and/or placed in the digital library; records created in NatureBib	• guidelines for data distribution and data discovery • guidelines for organizing project info (including a project file organizer and a template directory structure) • project information management SOP • SOP and template language for contracts and agreements (protected information, deliverables) • specifications for deliverables (includes reference to GIS specifications)
Design and testing	• protocol / methodology and SOPs • blank field forms and data dictionary • data design documentation • metadata – partially complete • Project file documentation – partially complete	• documents and supporting materials created and maintained in working project folders • final copies of protocols/SOPs sent to document archives and/or placed in the network digital library; records created in NatureBib	• naming conventions for files and data objects • data design standards • SOP - metadata procedures and specifications • guidelines for quality assurance and quality control • SOP – GPS data collection and processing
Implementation	• contracts and agreements • raw data forms, field notebooks, trip reports	• documents, databases, GPS rover files and GIS layers created and maintained in working project folders	• guidelines for managing third-party data • guidelines for labeling and storage of analytical samples

Project phase	Deliverables	Repository	Data management guidance
	• databases, GIS layers, GPS rover files, list of coordinates • biological specimens, species checklists, catalog of specimen data • photographs, illustrations, sound and video recordings • data certification report	• specimens to the park collection or other designated repository	and specimens • standards for voucher collection and cataloguing • photo management strategy • template language for contracts and agreements • naming conventions for files and data objects • SOP - metadata procedures and specifications • specifications for deliverables (including GIS specs) • guidelines for quality assurance and quality control • SOP – GPS data collection and processing
Product delivery and review	• annual report • trend analysis report • final reports - technical or general audiences • publications • metadata – completed • Project file documentation	• documents and supporting materials created and maintained in working project folders • final copies of reports sent to document archives and/or placed in the network digital library • permit investigator's annual reports to RPRS	• reporting guidelines and templates • specifications for deliverables (including GIS specs) • SOP - metadata procedures and specifications
Product integration	• finalized metadata posted on the internet • records in searchable corporate databases (e.g., NPSpecies, NatureBib) • products are secure and available	• NPS metadata clearinghouse • NatureBib • NPSpecies • network databases • all remaining materials moved to document archives and stored by calendar year	• integration of project data with corporate data • archival procedures for project records • sensitive info management strategy • NPSpecies and NatureBib guidelines • project information

Project phase	Deliverables	Repository	Data management guidance
			management SOP • *guidelines for data distribution and discovery*
Project evaluation and close out	• update record in project management database • completed checklists (data manager, permitting officer, project manager) • documentation of needed modifications	• network project management database • place completed checklist in document archives • maintain modifications documentation with related project documents	• *project information management SOP*

APPENDIX E: NPS PROJECT TRACKING SYSTEMS

PMIS

The **Project Management Information System** (**PMIS**) is a Service-wide NPS Intranet application designed to manage information about requests for project funding. It enables parks and NPS offices to submit project proposals to be reviewed, approved and prioritized at park units, regional directorates, and the Washington Office (WASO). PMIS contains historical information for project requests going back to FY99.

In response to a budget call for a particular NPS program for a specific fiscal year (FY), project proposals are submitted, reviewed, approved, prioritized and then formulated under an available funding source by utilizing PMIS. During formulation process for a budget call, a program manager at WASO or a budget officer at a regional directorate determines which project funding requests meet the eligibility criteria for the call to be considered as part of the NPS Budget for a specific FY.

RPRS

The **Research Permit and Reporting System** (RPRS) provides an electronic means to apply for research and collection permits. Park research coordinators use the system to issue and track all Scientific Research and Collecting Permits. Park research coordinators can use the system to perform the following major actions:

- Receive and organize electronic permit applications, proposals, and peer-reviews from applicants.
- Post and maintain the type of research the park is most interested in attracting.
- Post and maintain park-specific conditions applicable to every permit issued by the park.
- Post an information bulletin used to notify investigators of special conditions or events that could impact planned fieldwork (road closures, area closures, safety-related notices, etc.).
- Process and track permits and denied applications (including revoked or cancelled applications).
- Manage the park Investigator's Annual Report (IAR) database.
- Search the Service-wide IAR database.
- Search the Service-wide permit database to confirm currently active permits, previously approved studies conducted at other parks, and the reporting of annual accomplishments (submission of IARs) by investigators indicating that they have conducted previous studies in NPS units.

PEPC

The **Planning, Environment, and Public Comment (PEPC)** system is an on-line tool being developed that parks can use to help manage all stages of the compliance process. It includes a "public" side, where anyone can find out about activities going through compliance, and a "private" side for NPS information.

Appendix F: Data Stewardship Responsibilities

Role	Programmatic Responsibility	Data Stewardship Responsibilities
Project Crew Member	Collect, record, and verify data	Obtain training in data management for the project. Read and follow project protocols, study plans, and relevant NPS guidance. Communicate with Crew Leader, Project manager, and Data Manager. Record and verify observed or measured data values. Schedule and perform regular data transfer and backup. Review, verify, and correct field data. Assist with data and procedural documentation, especially deviations from the protocol or study plan.
Project Crew Leader	Supervise crew	Obtain training in data management for the project. Ensure crew members receive data management training and briefings. Read and follow all protocol, project, and relevant Network-level guidelines. Communicate with Crew Members, Project manager, and Data Manager. Ensure data are regularly transferred, backed up, verified, and entered into the appropriate NPS database(s). Assist with data and procedural documentation.
Data/GIS Specialist or Technician	Process and manage data	Obtain briefings about projects and related data to understand the geospatial and technical requirements and relevance. Communicate with other participants in the project to the extent necessary to accomplish assigned tasks. Perform assigned level of technical data management and/or GIS activities, including data entry, data conversion, and documentation. Work on overall data quality and stewardship with Project managers, Resource Specialists, and the Network Data Manager.
Information Technology/ Systems	Provide IT/IS support	Provide and maintain an information systems and technology foundation to support data management.

Role	Programmatic Responsibility	Data Stewardship Responsibilities
Specialist		Advise project participants about capabilities of hardware and software resources to support project and program objectives. Work with Database Manager to resolve hardware and software issues relating to database functions and availability.
Project manager	Oversee and direct project operations	Ensure Crew Leader receives pertinent training and briefings. Communicate with Crew Leader, Data Manager, and I&M Network Coordinator. Complete project documentation describing the who, what, where, when, why and how of a project. Develop, document and implement standard procedures for field data collection and data handling. Enact and supervise quality assurance and quality control measures for the project. Supervise and certify all field operations, including staff training, equipment calibration, species identification, and data collection. Supervise or perform data entry, verification and validation. Maintain concise explanatory documentation of all deviations from standard procedures. Ensure documentation of important details of each field data collection period. Maintain hard copies of data forms and send original data forms to archive on a regular basis. Work with program coordinators to identify analysis and reporting mechanisms, and to establish a schedule for regular project milestones such as data collection periods, data processing target dates, and reporting deadlines. Produce regular summary reports and conduct periodic trend analysis of data, store the resulting reports, and make them available to users. Act as the main point of contact concerning data content. **The project manager works closely with the data manager to:** Develop quality assurance and quality control procedures specific to project operations.

Role	Programmatic Responsibility	Data Stewardship Responsibilities
		Identify training needs for staff related to data management philosophy, database software use, quality control procedures, etc. Coordinate changes to the field data forms and the user interface for the project database. Fully document and maintain master data. Identify sensitive information that requires special consideration prior to distribution. Manage the archival process to ensure regular archival of project documentation, original field data, databases, reports and summaries, and other products from the project. Define how project data will be transformed from raw data into meaningful information and create data summary procedures to automate and standardize this process. Identify and prioritize legacy data for conversion; convert priority data sets to a modern format. Increase the interpretability and accessibility of existing natural resource information. **Note:** The Project manager is often a resource specialist, in which case the associated responsibilities for data authority apply (see resource specialist role). A Project manager without the required background to act as an authority for the data will consult with and involve the appropriate Resource Specialists.
Resource Specialist	Understand the project and make decisions about the data	Understand the objectives of the project, the resulting data, and their scientific and management relevance. Guide development of an Information Needs Assessment based on the objectives of the project. Make decisions about data with regard to validity, utility, sensitivity, and availability. Describe, publish, release, and discuss the data and associated information products. **Note:** The Resource Specialist serving as a Project manager is also responsible for the duties listed with that role.
GIS Manager	Support park management objectives with GIS and resource information	Coordinate and integrate local GIS and resource information management with Network, Regional, and National standards and guidelines. **The GIS specialists will work in collaboration with project managers to:**

Role	Programmatic Responsibility	Data Stewardship Responsibilities
	management	Determine the GIS data and analysis needs for the project. Develop procedures for field collection of spatial data including the use of GPS and other spatial data collection techniques. Display, analyze, and create maps from spatial data to meet project objectives. Properly document data in compliance with spatial metadata standards. **GIS specialists will also work directly with data managers to:** Design databases and other applications for the network. Create relationships between GIS and non-spatial data and create database and GIS applications to facilitate the integration and analysis of both spatial and non-spatial data. Establish and implement procedures to protect sensitive spatial data according to project needs. Develop and maintain an infrastructure for metadata creation and maintenance. Ensure that project metadata are created and comply with national and agency standards.
Network Data Manager	Ensure inventory and monitoring data are organized, useful, compliant, safe, and available	Assist in developing and implementing procedures to ensure that I&M data collected by NPS staff, cooperators, researchers and others are entered, quality-checked, analyzed, reported, archived, documented, cataloged, and made available to others for management decision-making, research, and education. Provide guidance and support, to the extent possible, to extend Network standards and procedures to studies and data funded by park base and other funding sources to promote integration and availability of datasets. Provide overall Network planning, training, and operational support for the awareness, coordination, integration of data and information management activities, including people, information needs, data, software, and hardware. Serve as Point of Contact for National Park Service database applications (NPSpecies, NatureBib, Dataset Catalog) Coordinate internal and external data management activities. Assign and enforce data stewardship responsibilities.

Role	Programmatic Responsibility	Data Stewardship Responsibilities
		Review and approve all data acquisition plans, hardcopy and electronic field forms, and data dictionaries. Participate in development of Information Needs Assessments. Communicate with Crew Leader, Project manager, I&M Network Coordinator, and Park GIS/Data Management office. Develop and maintain overall Network and individual Vital Sign data management operating guidelines and relationship to national standards and procedures. Develop and maintain the infrastructure for metadata creation, project documentation, and project data management. Create and maintain project databases in accordance with best practices and current program standards. Provide training in the theory and practice of data management tailored to the needs of project personnel. Develop ways to improve the accessibility and transparency of digital data. Establish and implement procedures to protect sensitive data according to project needs. Collaborate with GIS Specialists to integrate tabular data with geospatial data in a GIS system in a manner that meets project objectives. **Data managers will also work closely with the project manager to:** Define the scope of the project data and create a data structure that meets project needs. Become familiar with how the data are collected, handled, and used. Review quality control and quality assurance aspects of project protocols and standard procedure documentation. Identify elements that can be built into the database structure to facilitate quality control, such as required fields, range limits, pick-lists and conditional validation rules. Create a user interface that streamlines the process of data entry, review, validation, and summarization that is consistent

Role	Programmatic Responsibility	Data Stewardship Responsibilities
		with the capabilities of the project staff. Develop automated database procedures to improve the efficiency of the data summarization and reporting process. Make sure that project documentation is complete, complies with metadata requirements, and enhances the interpretability and longevity of the project data. Ensure regular archival of project materials. Inform project staff of changes and advances in data management practices. Additional examples of the duties and responsibilities of the network data managers are listed in I&M Program Vision and Organizational Framework document "Network Data Manager Overview of Responsibilities". **NOTE:** Data Managers with Prototype Monitoring Programs have the same basic duties and responsibilities as the network data managers but also are responsible for mentoring and training others and developing and testing new approaches to data analysis, synthesis, and reporting of monitoring results.
Database Manager	Know and use databases and applications	Install, maintain, and support specific database software applications and NPS database applications. Work with Information Technology Specialists to resolve hardware and software issues.
Curator	Oversee all aspects of the acquisition, documentation, preservation, and use of park collections	Know park natural resource collections Conduct accessioning, cataloging, legal, and other documentation of collections Manage collections databases Recognize objects needing conservation treatment Recommend and refer treatment to the appropriate facility Work with Network Data Manager to acquire and process data related to natural resource collections
Statistician or Biometrician	Analyze data and present information	Work with the Network Ecologist to analyze and report data according to established protocols. Work with the Network Data Manager to acquire and process raw data from databases and store derived data and information after analysis
Network Ecologist	Integrate science in network	Ensure useful data are collected and managed by integrating natural resource science in network activities and products, including objective setting, sample design, data analysis,

Role	Programmatic Responsibility	Data Stewardship Responsibilities
	activities	synthesis, and reporting. Assist with development and modification of monitoring protocols and inventory study plans. Work with the Network Data Manager to incorporate data management in monitoring protocols. Participate in the development of Information Needs Assessments based on the objectives of the project. Guide and/or perform statistical and other analyses of network data. Contribute to the synthesis and reporting of data and information. Provide guidance and support, to the extent possible, to extend Network standards and procedures to studies and data funded by park base and other funding sources to promote integration and availability of datasets.
Network Coordinator	Coordinate all network activities	Ensure programmatic data and information management requirements are met as part of overall Network business. Communicate with Network staff, park staff at all levels, and other appropriate audiences to support and emphasize data management as a critical aspect of network business Work with Network Data Manager regarding data management policy and guidelines, budget, staffing, and training. Hold Network staff accountable for responsibilities involving data management.
I&M Data Manager (National Level)	Provide Service-wide database availability and support	Provide services to receive, convert, store, and archive data in Service-wide databases. Work with Network Data Manager to resolve local issues involving the access and use of inventory and monitoring databases. Provide training where possible. Design and maintain standardized, master databases for Service-wide planning, decision-making, and accountability (e.g., NPSpecies, NatureBib, Dataset Catalog, Database Template, GIS tools). Collaborate with networks to help develop overall data management vision and approach, and continual improvement of specific tools.

Role	Programmatic Responsibility	Data Stewardship Responsibilities
		Coordinate establishment of standards for naming conventions and content of data management plans and monitoring protocols. Promote collaboration and integration with other divisions and programs including the GIS community, fire program, air resources, water resources, geologic resources, etc. Facilitate coordination and collaboration among the parks and networks by providing examples of good database designs with flexibility to allow adjustments for different situations.
Other End Users	Use and apply Network services and products	These 'information consumers' include park managers and superintendents, researchers, staff from other agencies, and the public. End users at all levels are generally responsible for providing necessary and requested feedback, review, and comments on various products in order to sustain the continuous improvement of network operations and services. End users are responsible for the appropriate use and application of data and derived products.

Appendix G: I&M Program Network Data Manager Position Description

OVERVIEW	***Foster the success of the I&M program through data stewardship; ensure quality data sets throughout the network, integrate data and coordinate and communicate openly with all data managers throughout the region, and contribute where appropriate to the national I&M efforts.***

	LEVEL OF RESPONSIBILITY	
DUTY	*RESPONSIBLE FOR* ***(core duty)***	*INVOLVED IN/AWARE OF* ***(secondary duty)***
Design Databases	• Understand and follow conventions of the I&M Database Template • Ensure "good" relational database modeling practices are followed for all I&M projects across the network • Ensure synchronization of tabular and spatial data models for all I&M projects across the network • Coordinate efforts between similar data models within the network • Ensure that appropriate design documentation has been created, such as data dictionaries and model diagrams • Ensure database designs are compatible with field collection needs	• Assist with non-I&M projects to ensure quality database designs • Coordinate efforts between similar data models across all networks in region • Contribute to national efforts of designing cohesive datasets

DUTY	***LEVEL OF RESPONSIBILITY***	
	RESPONSIBLE FOR ***(core duty)***	*INVOLVED IN/AWARE OF* ***(secondary duty)***
Implement Databases / Maintain Datasets	• Ensure appropriate database software is selected for all I&M projects across network • Ensure suitable cataloging, storage, and accessibility of digital data • Develop archiving procedures and ensure their practice • Evaluate legacy data for usability • Develop and implement QA/QC procedures and data validation tools	• Coordinate efforts between similar datasets across all networks in the region
Populate Data Sets	• Ensure reliability of "location" data collection methods for accurate spatial data • Ensure data integrity as data sets are converted to electronic format • Provide programming skills for data entry forms and other input methods • Ensure FDGC-compliant metadata for spatial datasets • Generate spatial datasets from tabular data • Analyze data to identify potential data anomalies • Ensure population of national datasets such as Dataset Catalog, NPSpecies, ANCS+, and NPBib • Ensure data allows for appropriate setup of ArcView to Access Links	• Contribute to national efforts of collecting cohesive datasets

DUTY	***LEVEL OF RESPONSIBILITY***	
	RESPONSIBLE FOR ***(core duty)***	*INVOLVED IN/AWARE OF* ***(secondary duty)***
Provide Data Sets/Products	• Assist others in the understanding and analysis of collected data • Format data sets as required by others for reporting or analysis tools • Develop desktop (distributable) and web-based applications for data entry, viewing, and reporting • Develop GIS products for data analysis and dissemination	• Assist statisticians with developing data formats to fit statistical models or analysis packages • Assist with creation of ArcView to Access links • Assist with creation of graphics presentations

NON-DUTIES	***LEVEL OF RESPONSIBILITY***
	NOT RESPONSIBLE FOR
	• Park GIS and database solutions outside of the I&M Program; should assist when skills and expertise are required, but should not be solely responsible for implementing projects at the park level • System or network administration and support, including hardware and software issues • General data entry; should not be responsible for keying data in from field forms • Metadata and documentation generation

APPENDIX H. METADATA GENERATION TOOLS

NPS MetaData Tools & Editor

This tool was released in 2005, and consolidates some of the metadata creation tasks used within the NPS into one tool. The Metadata Tools and Editor (MTE) can be used to build metadata that conforms to the MPS Metadata Profile, including FGDC and NBII Profile elements. In addition, the MTE is set up to include many NPS defaults and picklist elements, greatly facilitating Metadata development for NPS users. Functionally, the MTE is a stylesheet editor which operates on XML. This tool can be used from a standalone application, or as a stylesheet addition to the metadata tools in ArcCatalog (see below). Revisons and enhancements to this application continue to be developed, with the latest version release in July of 2006.

ArcCatalog

ArcCatalog is a management tool for GIS files contained within the ESRI ArcGIS Desktop suite of applications. With ArcCatalog, users can browse, manage, create, and organize tabular and GIS data. In addition, ArcCatalog comes with support for several popular metadata standards that allow one to create, edit, and view information about the data. There are editors to enter metadata, a storage schema, and style sheets to view the data. With ArcCatalog users can view GIS data holdings, preview geographic information, view and edit metadata, work with tables, and define the schema structure for GIS data layers. Metadata within ArcCatalog is stored exclusively as Extensible Markup Language (XML) files. The NPS Integrated Metadata System Plan recommends ArcCatalog for gathering GIS-integrated geospatial metadata.

Dataset Catalog

Dataset Catalog is a tool for cataloging abbreviated metadata on geospatial and biological data sets pertaining to park(s) and/or a network. It provides parks and/or networks a means whereby they can inventory, organize, and maintain information about data set holdings locally. While Dataset Catalog is not intended to be an exhaustive metadata listing, it does assist parks and networks in beginning to meet the mandates of EO 12906. Dataset Catalog ability to export records in Extensible Markup Language (XML), and this allows it to be used as a 'first tool' for crafting metadata records. The I&M Program recommends that all relevant datasets at I&M parks and networks be cataloged in at least simple Dataset Catalog format.

Metadata Parser

The MetaParser (mp) program is used to validate metadata records by checking the syntax against the CSDGM and to generate compliant output files for posting to clearinghouses. It generates a textual report indicating errors in the metadata, primarily in the structure, but also in the values of some of the scalar elements where values are restricted by the standard.

Appendix I. Project Archive Checklist

National Park Service
Inventory & Monitoring Program
Great Lakes Network

Project Check List

Place a copy of this completed form in the hardcopy folder of the project. Store an electronic copy of this document with appropriate computer files.

Project Name: _____
I&M Project Number: _____

Project Personnel: _____

Project Contact (email/phone): _____

Date (form completed): _____

Location of Materials:
Physical files: _____
Electronic files: _____

Check List:

√	Project Elements	File Type*	Entered Redisc.	NPS Archive†	Comments/Specific Location (of either physical or digital items):
	Final Proposal				
	Study Plan / Protocols				
	Permit				
	Compliance Docs				
	Field Forms†				
	Field Notes†				
	Tabular Data (databases, spreadsheets, etc.)				File name and location:

	Reports				
	Physical Maps†				
	GIS Files				
	Photographs				
	Other Files (presentations, posters, graphics, related literature, etc.)				
	Digital backup of project†				(circle as appropriate): DVD CD Other:
	HD backup (shared drive)				
	Reports to Park Libraries				
	NatureBib Entry				
	NPSpecies Entry				
	Metadata (digital and hardcopy)				
	Closeout of Project/Permits				

*File Type: PF = Project Folder (hardcopy in project folder), RF = Park Resource Library Folder (hardcopy in library folder), PE = Project Electronic Folder, O = Other (specify)

Notes:

APPENDIX J: GLKN SPATIAL LAYER RESOURCES

GIS Layer	APIS	GRPO	INDU	ISRO	MISS	PIRO	SACN	SLBE	VOYA
AIR PHOTO IMAGERY									
DOQQ	1992 - all /1998 (Bayfield only)	1998	1998	1998	1992/2003	1998	1992/1998	1998	1992 (Most of park missing)
DOQQ_SID	X	X	X	X	X	X		X	X
DRG	X	X	X	X	X	X	X	X	X
DRG_SID	X	X	X	X	X	X	X	X	X
Recent Air Photos	1:12,000 CIR (not ortho) 1:24,000, Ortho 0.30m	1:8,000 Ortho 0.15m	1:12,000 0.20m Ortho			1:12,000 CIR, 0.20m ortho			
Historical Air Photos	1938		1939		1937	1939		1938	1927
SATELLITE IMAGERY									
Landsat	16	Same scenes as APIS, ISRO	13	3 different scenes 14, 4, 11	17	10	3 different scenes 16, 10, 16	16	13
IKONOS	IKONOS mainland summer 2004		IKONOS Summer 2004			IKONOS Summer 2004		IKONOS Summer 2004	IKONOS Summer 2003
DEM10m	X	X	X	X	X	X	X	X	X
DEM30m	X	X	X	X	X	X	X	X	X
DEM_other				2m LiDAR					
BASE CARTOGRAPHY									
Hypsography	Bayfield Co. only								
Hydrography	NHD, 12k from soil survey	NHD 24k	NHD 24k	NHD 24k	NHD 24k	NHD, 12k in progress	NHD 24k	NHD 24k	NHD 24k
Wetlands	WWI	NWI	NWI	NWI	NWI	NWI	NWI	WWI	NWI
Watershed	8 digit HUC	8 digit HUC	8 digit HUC	8 digit HUC	8 digit HUC	8 digit HUC	8 digit HUC	8 digit HUC	8 digit HUC
Geology		X		X	x (Landforms)				
Soils	in progress	in progress	X	in progress	X	in progress	in progress	X	
Fire			X						
Park Boundary	X	X	X	X	X	X	X	X	X
Ownership	X						Partially complete		
Roads	2003 TIGER files	2003 TIGER files	2003 TIGER files	2003 TIGER files	2003 TIGER files	2003 TIGER files	2003 TIGER files	2003 TIGER files	2003 TIGER files
Trails	X	X	X	X	X	X	X	X	X
Campsites	X					X			X
Buildings		X							
Vegetation	NVCS in progress		NVCS in progress	NVCS complete	MLCCS (1999)				NVCS complete

CIR	Color Infra Red
DEM	Digital Elevation Model
HUC	Hydrologic Unit Code
MLCCS	Minnesota Land Cover Classification System
NHD	National Hydrography Data
NVCS	National Vegetation Classification System
NWI	Naitonal Wetland Inventory
TIGER	Topologically Integrated Geographic Encoding and Referencing system
WWI	Wisconsin Wetland Inventory

APPENDIX K: GLKN GPS FIELD DATA COLLECTION GUIDE

Version 1.0 - September, 2006

Introduction

Over the past decade, new tools have been developed to help researchers collect information about spatial locations while in the field. Global Positioning System (GPS) is one such tool. GPS is currently a constellation of 28 US Department of Defense satellites (as of 2006) orbiting 11,000 miles above the Earth, making a complete orbit approximately every 12 hours, and transmitting signals to Earth at precisely the same time. The position and time information transmitted by these satellites is used by a GPS receiver to triangulate a location coordinate on the earth using three or more satellites.

This guide supplements other sources of information available to GLKN GPS users:

1. National Park Service *Field Data Collection with Global Positioning Systems Standard Operating Procedures and Guidelines, 02/10/2004.* (http://www.nps.gov/gis/data_info/standards.html)
2. Individual project specifications or protocols that should contain additional detailed standard operating procedures for GPS data collection.
3. Manufacturer's user guides and operations manuals for GPS hardware and software.

This guide addresses GPS instruments, instrument settings, field operation, data collection, data processing, and a standardized method for acquiring and managing location data. It contains the procedures and considerations that GPS users should follow when collecting geospatial data or navigating to locations in Great Lakes Network Network (GLKN) parks.

Role of GPS in GLKN Data Management

Data collected using GPS-enabled equipment can represent all or part of the acquisition stage of a data management cycle that also includes several other stages (see Section 5.4 in GLKN Data Management Plan). The process and methodology used for acquisition planning, data collecting, and post-processing incorporate several aspects of data management, including quality assurance, data storage and organization, and data stewardship. To promote data quality and simplify data management, the Great Lakes Network expects to use electronic data logging equipment for data acquisition. However, parallel or complementary use of hand written data sheets and field notes will remain important for some data collection activities.

Types of GPS units

At the most basic level, GPS equipment consists of a GPS antenna and the associated signal processing circuitry. The antenna can be a standalone device, incorporated into a handheld unit, or integrated into a larger electronic device, such as a personal data assistant (PDA), data logger, or portable computer. As technology evolves, the Great Lakes Network will continually strive to use equipment which maximizes spatial

accuracy, reduces hardware weight and user fatigue, and reduces database development, data manipulation and transformation.

There are three major types of GPS units based on the level of accuracy to which spatial data can be collected. Survey-grade GPS units are used for surveying tasks that require very high accuracy (1 cm or less). Mapping-grade units can map features from sub-meter to better than 5-meter accuracy, employing differential correction. Recreational-grade GPS units are sold primarily for outdoor sports and recreational activities, and can provide accuracies ranging from 5 – 30 meters. Mapping-grade or recreational grade units are suitable for most natural resource-related data collection requirements. Figure 1 shows some of the major differences between these two types.

Deciding which type of unit to use is an essential part of project planning, and depends on the end product needed. Mapping-grade GPS units are recommended for most GLKN field work; however, for some projects recreational-grade units can meet a project's accuracy requirements and reduce the cost of field operations. Most natural resource monitoring consists of repeated visits to the same location, which can often be identified in the field by a marker or other characteristic. In this case a mapping grade unit is probably desirable to establish a very accurate location at the start of a study; whereas a recreation-grade GPS receiver integrated into a field data collection PDA could be used on subsequent visits. In this case the position fix from the recreational-grade unit provides confirmation of both location and time. The choice of GPS unit should be made by the project manager after consulting with the GLKN data management and GIS staff.

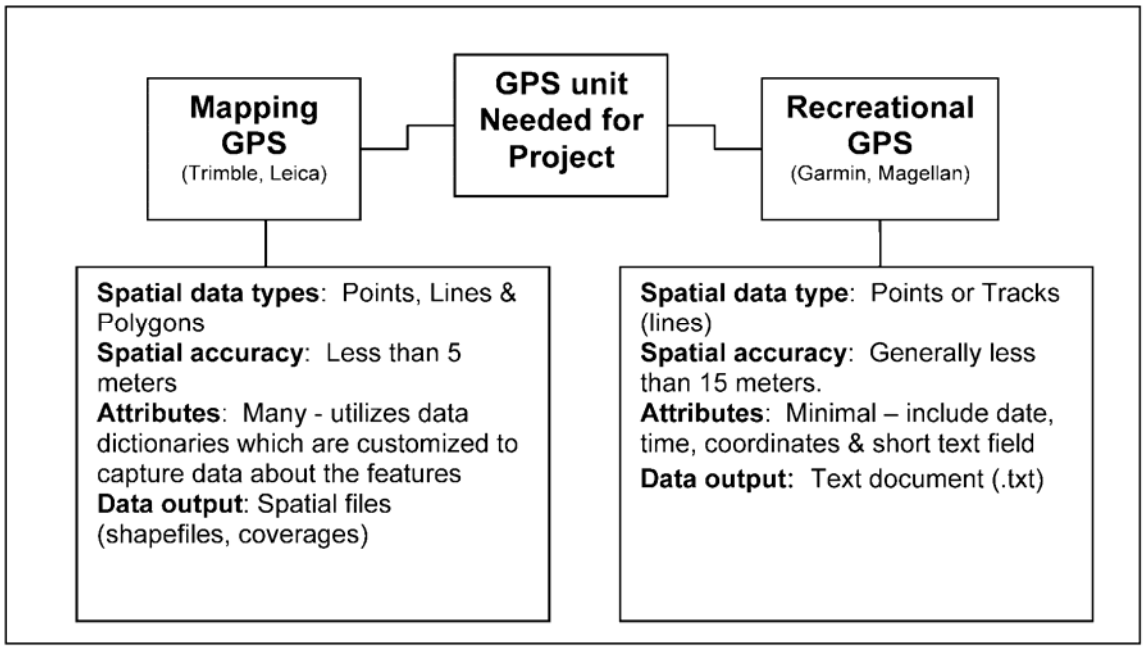


Figure 1. Differences in accuracy and metadata capabilities between types of GPS units

All resulting GIS data layers need to meet or exceed National Map Accuracy Standards for a 1:24,000 product (NPS GIS Data Standards, 2002 http://www.nps.gov/gis/data_info/standards.html). Table 1 provides the allowable horizontal accuracy for some common scales.

Table 1. Map scales and allowable error

Scale	Allowable Error
1:40,000	33.8 meters (111 feet)
1:31,680	16.1 meters (53 feet)
1:24,000	12.2 meters (40 feet)
1:20,000	10.1 meters (33 feet)
1:12,000	6.1 meters (20 feet)
1:9,600	4.9 meters (16 feet)
1:4,800	2.4 meters (8 feet)
1:2,400	1.2 meters (4 feet)
1:1,200	0.6 meters (2 feet)

Using Mapping-grade GPS Units

Mapping-grade GPS units provide the user with a variety of tools for field data collection. GLKN encourages the use of these units for most projects where location information will be integrated into GIS data layers. These units can acquire spatial data related to points, lines and polygons along with associated, user defined, tabular attributes. Careful forethought and advanced planning are required to take advantage of these capabilities before data collection begins.

Data dictionaries

Mapping-grade GPS units are capable of using data dictionaries. Data dictionaries define the structure and rules to store attribute information about the feature being mapped and are customized for each project. GLKN data management staff should be directly involved in the creation of data dictionaries. Basic steps include:

1. Identify the features to be mapped. These features are real world physical locations of objects which are categorized as point, line or polygon features.
2. Identify the data to be collected about each feature while in the field and create a data dictionary. Part of this process involves assigning a unique identifier to each feature. (For example, if a survey plot is mapped as both a point and a polygon, one feature should be named plotname_poly and the other plotname_pt.)
3. Implement and test the data dictionary. A complete trial run for newly-created data dictionaries should be completed before beginning field work. Corrections and refinements are inevitable after such a trial.

GPS Settings

Positional accuracy of GPS data can be affected by several factors which can be monitored and recorded with mapping-grade GPS units. Table 2 lists these factors, their definitions, and the standard settings for GLKN field work. If real-time differential correction (DGPS), WAAS or SBAS, is available it should be enabled and used. All spatial data collected shall be analyzed for spatial accuracy and must meet or exceed National Map Accuracy Standards (see Table 1 and http://mapping.usgs.gov/standards/). Table 3 indicates the coordinate system settings for data collection in GLKN parks.

Table 2. GPS Receiver Settings and Minimum Standards

Name	Definition	GLKN Standard
Almanac	File containing estimated position of satellites, time corrections, and atmospheric delay parameters.	Acquired automatically by GPS unit or from online sources within 10 days prior to GPS field work.
Altitude reference	Ellipsoid model	Height above Ellipsoid (HAE) (preferred) or Mean Sea Level: if MSL is used, indicate Geoid Model
Antenna Height	GPS antenna height above the ground.	Variable, usually 1.0 meters for handheld and 1.5 m for backpack.
Datum	Geodetic model designed to fit a point on the earth's surface to an ellipsoid.	NAD 83 (preferred) WGS 84 (GPS default, as fallback)
Elevation mask	The minimum angle above the horizon at which a GPS receiver will track a satellite.	15 degrees
Feature types	Geometry of spatial data	GIS native formats; point, line and polygon are preferred.
Logging interval	Time interval between the recording of individual GPS fixes.	Points: 1 second Lines and Polygons: 5 seconds, but 1 second in some circumstances.
Minimum fixes for point positions	Number of GPS fixes that are used to calculate a single position for a point feature.	50 fixes
Mode	2 dimensional for horizontal positions and 3 dimensional for an elevation position.	3-dimensional (4 satellite minimum)
PDOP Mask	Positional Dilution of Precision, a GPS quality estimate based on satellite geometry.	6.0 or less.
Satellite vehicles	Number of satellites used for position fixes.	4 minimum
SNR Mask	Signal-to-Noise ratio is a measure of the satellite signal relative to background noise.	4.0 minimum, 6.0 or greater preferred.
Unit of Measure	Linear unit of measure	Meter (metric)

Table 3. Coordinate system settings for Great Lakes Network parks.

Park	UTM Zone	Datum
APIS	15	NAD 1983 (CONUS)
GRPO	16	NAD 1983 (CONUS)
INDU	16	NAD 1983 (CONUS)
ISRO	16	NAD 1983 (CONUS)
MISS	15	NAD 1983 (CONUS)
PIRO	16	NAD 1983 (CONUS)
SACN	15	NAD 1983 (CONUS)
SLBE	16	NAD 1983 (CONUS)
VOYA	15	NAD 1983 (CONUS)

Universal Transverse Mercator (UTM) should be the preferred choice for a native projection system of GPS data. The proper UTM zone of GLKN parks is found in Table 3. If Latitude-Longitude is used instead of UTM, decimal degrees is the proper format, not degrees, minutes and seconds.

Before beginning data collection, the GLKN data management staff or the GPS user should complete mission planning tasks. If high accuracy fixes are desired and there is some flexibility in scheduling of a field data collection mission, the user should complete a satellite survey to determine the best timing of the mission, usually when the most satellites are visible and in the best geometry. Software, such as Trimble's Quick Plan (also included in Pathfinder Office) can be used to look for time windows that should offer the lowest PDOP readings, and thus the highest positional accuracy.

If a project requires navigation to preset locations or waypoints, this information must be pre-loaded onto the GPS hardware before starting a field data collection mission. Having printed copies of topographic maps with waypoints marked is a good backup and can maximize field time and efficiency when navigating between waypoints and sites.

Many mapping-grade GPS units have the capability of storing and displaying background maps or GIS layers, which can be very helpful when navigating in the field. The GLKN data management staff can assist in preparing these background layers and loading them onto the GPS hardware.

GPS units create files to store data during a field session using a prefix and date-time stamp as file names:

RMMDDHHx
R – Unit Prefix
MM – Month
DD – Day
HH – Hour
X – a, b, c, etc., the order files are created within an hour

If multiple GPS units are used for a project, a unique prefix (letter) should be assigned to each unit, which will ensure that downloaded files for each unit contain a unique identifier within the filename. For example, with three GPS units, the unique letters for the units could be N, G, and A. Those letters would serve as a prefix for the file n(e.g., N102715A, G102715A and A102715A would indicate units N, G, and A, October27, 15 hour, A first in hour). The exact file naming standard for a protocol must be spelled out in its data management or GPS documentation.

Each user should be familiar with the capabilities of the GPS hardware and field computers. All units have limited battery and memory resources, and a balance needs to be reached that will allow the most efficient use of the resources. Some units have exchangeable batteries, and almost all can be charged from a 12 volt vehicle power port.

Data Collection

Data collection should be performed using an approved data dictionary or database for the protocol. Users should be mindful of the following concepts:

- If using a GPS unit that gives an approximate 5 meter horizontal accuracy, the user cannot accurately map anything as a polygon that is less than 10 meters in width or diameter. Such objects must be captured as point features.

- If a GPS user is collecting a line or polygon feature and then stops moving, the GPS unit will continue to collect data (Figure 2, ex. 1 and 2). Users need to be familiar with the *Pause/Resume* toggle key and use it liberally.
- Another way to avoid errors is to collect point features that represent the beginning and end points of a line transect (Figure 2, ex. 3 and 4). Having these reference point locations will mean easier editing of any zig-zagging line features.

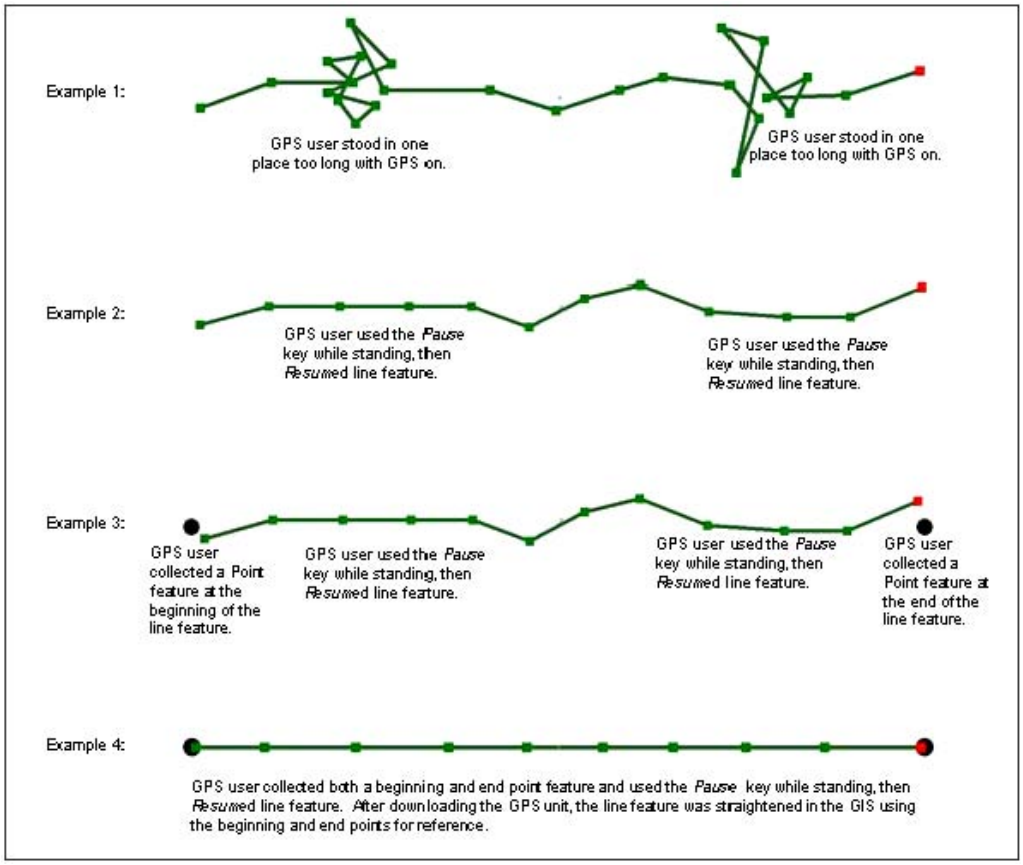


Figure 2. Examples of GPS point feature data collection

Mapping-grade GPS units have additional features that aid in data collection. These include:

- *Nested Features* – Allows user to collect a point while collecting a line or polygon feature. For example, while surveying potential amphibian habitat along a stream (line feature), the user can pause the line feature, take a point a specific observation, then resume the line feature collection.
- *Offset Feature* – Allows user to collect a feature when topography is such that getting next to or over the feature is impossible. For example, a GPS point could be collected for a hanging garden while standing to the side of the actual garden.
- *Between-feature Positions* – the GPS unit collects positions without any feature or attribute data. This feature is useful for tracking areas traveled during a day.

For example, while surveying for presence or absence of invasive plants, a user could collect data on the area surveyed in addition to specific plants found.

Data Processing:

When data collection is complete for the day or field visit, data are downloaded from the GPS unit to a computer. For Trimble GPS units, the proprietary software Pathfinder Office (or the GPS Analyst extension) is used to download, differentially correct and then export the data to a GIS format. Post processing differential correction is procedure to improve upon raw, or real-time DGPS, GPS positions using base station data. Base stations consist of a GPS antenna and receiver positioned at a known location specifically to collect data from satellites. The distance between the base station and the remote GPS receiver should be kept to a minimum, preferably less than 100 miles.

Once the data are differentially corrected, they can be verified and edited. Unintentional features can be deleted and attributes can be reviewed. The last step is exporting the data set to GIS (ArcGIS or ArcView). Additional data attributes can be included in the data exports. Data attributes recommended by GLKN are listed in Table 4.

Table 4. Recommended fields to be exported in addition to GPS features

All Features	Point Features	Line Features	Area Features
PDOP	Height	Length (2D)	Area (2D)
Correction Status	Position	Length (3D)	Perimeter (2D)
Receiver Type			Perimeter (3D)
Date Recorded			
Data file name			
Total positions			
Data Dictionary name			

Managing the incoming GPS data can be a challenge, especially if there are multiple units per project. Common practices used by GLKN include:

- Download all data to a computer or network drive that is regularly backed up.
- Keep GPS data and GIS data separate.
- Directories and files names should not contain non-alpha-numeric characters and/or spaces.
- Keep GPS data in well-organized directories (see GLKN Data Management Plan).

At the end of a project, all data and background files should be removed from the GPS unit. Data files should not be left on a unit if they have been properly downloaded and verified.

Additional information can be found at **http://www.nps.gov/gis/gps/gps4gis/** , which describes the steps outlined here in greater detail.

Using Recreational-grade GPS Units

Recreational-grade GPS units can be used to acquire location information (generally points) when high spatial accuracy is not essential to the project. Recreational GPS units do not have data dictionaries for storing attribute information with the point location, nor do most of them display or store significant metadata about the positional accuracy.

Planning

If a recreational-grade GPS meets the criteria of the project, the unit chosen must have the capability of downloading collected data to a personal computer. This is usually accomplished with a parallel or USB cable connection.

Much of the data collected by GPS will eventually reside in a relational database. Each GPS feature collected should contain a unique identifier that relates the feature to an associated record in a database. Since recreational GPS units have only one text field for input, careful consideration should be given to the use of this field and the design of unique identifiers. GLKN data management and GIS staff can assist in creating unique IDs on a project by project basis.

Although many recreational-grade units do allow a background map to be displayed, this option is usually restricted to a small number of topographic or highway navigation maps provided by the hardware vendor.

Data Collection

Location data are captured by recreational-grade GPS units as *waypoints*. When taking a waypoint, enter the Location ID in the text field provided. It is also good practice to collect reference points at regular intervals. Reference point positions should be taken at known locations (e.g., trailheads, parking lots, stream confluences) which can later be used in GIS to QC the accuracy of waypoint data.

If navigation to preset waypoints is applicable to a project, they must be loaded onto the GPS unit before departure to the field. Printed topographic maps of the waypoint locations can also be used to maximize field time and efficiently navigate between waypoints.

Data Processing

GPS units should be downloaded once a day or after each field session. The DNR Garmin freeware product (**http://www.dnr.state.mn.us/mis/gis/tools/arcview/index.html**) can be used to download data from the GPS unit. Data should be downloaded both as a text file and a shapefile. Each file name should include the download date. Although most recreational-grade GPS units collect and display positions in degrees, minutes and seconds by default, the waypoint locations should be exported in UTM projection or in decimal degrees format. Points should be checked for spatial accuracy and obvious errors. Subsequent downloads should be error-checked in the same manner. When data collection is finished, all files should be compiled into one spatial file, and along with the raw downloads, saved to the appropriate location on Great Lakes Network servers.

Metadata

Regardless of the type of GPS unit used to collect data, all resulting GIS datasets need to have information documenting how the GPS data were collected. NPS requires that FGDC (Federal Geographic Data Committee, **www.fgdc.gov/index.html**) compliant metadata be written for all geospatial layers created (Executive Order 12906).

The Great Lakes Network recommends formal metadata be written by the data collectors as they are the ones familiar with the project and resulting data. However, Network data management and GIS staff are usually the ones documenting someone else's work. At a minimum, the following details should be documented to facilitate final FGDC metadata:

- Name of project
- Name(s) of data collectors
- EHE/EPE or maximum PDOP (using 4 satellites)
- Coordinate system (projection, datum & zone)
- Type (or types) of GPS units used
- The range of field collection dates
- Name of Base Station(s) used for differential correction
- Name and version of software used for downloading
- Any major editing performed on the raw data (e.g. moving of points)
- All versions of data dictionaries should be saved

Credits

Portions of this guide were adapted from concepts and material developed by the Northern Colorado Plateau Network.

APPENDIX L: ACRONYMS

ANCS+	Automated National Catalog System
CDM	Conceptual Data Model
CSDGM	Content Standards for Digital Geospatial Metadata
DMP	Data Management Plan
EPA	U.S. Environmental Protection Agency
ERD	Entity Relationship Diagram (for databases)
ESRI	Environmental Systems Research institute
FGDC	Federal Geographic Data Committee
FOIA	Freedom of Information Act
FTP	File Transfer Protocol
GIS	Geographic Information System
GLKN	Great Lakes Inventory & Monitoring Network
GPS	Geographic Positioning System
HTML	Hypertext Markup Language
I&M	Inventory & Monitoring (Program)
IMS	Internet Mapping Server
IT	Information Technology
ITIS	Integrated Taxonomic Information System
JPEG	Joint Photographic Experts Group
LAN	Local Area Network
LDM	Logical Data Model
LiDAR	Light Detection and Ranging
MOU	Memorandum of Understanding
mp	Metadata Parser
MS	Microsoft – change in text to Microsoft
MWR (MWRO)	Midwest Region(al Office)
NBII	National Biological Inventory Infrastructure
NOAA	National Oceanic and Atmospheric Administration
NPS	National Park Service
NRDT	Natural Resource Database Template
NR-GIS	Natural Resource – Geographic Information System
NRTS	Natural Resource Table Standards
OMB	Office of Management and Budget
PC	Personal Computer
PDA	Portable Digital Assistant
PDF	Adobe Portable Document Format
PDM	Physical Data Model
QA/QC	Quality Assurance/Quality Control
RDMS/RDBMS	Relational Database Management System
RFID	Radio Frequency Identification
SDE	Spatial Database Engine
SOP	Standard Operating Procedure
SQL	Structured Query Language
TIFF	Tagged Image File Format
USGS	U.S. Geological Survey
WAN	Wide Area Network
WASO	Washington Support Office
XML	Extensible Markup Language

APPENDIX M: GLOSSARY

ArcIMS, ArcSDE, (Arc Internet Map Server and Arc Spatial Database Engine) are ESRI GIS software applications that provide tools for managing large, complex data sets within a relational database management system – ArcSDE, and serving the data over the internet - ArcIMS.

Conceptual data model is a detailed model that shows the overall structure of organizational data while being independent of any database management system or other implementation considerations. It is typically produced as the first modeling step in system design, and it is frequently a precursor to the logical data model. Conceptual data models are often presented as entity relationship diagrams.

Data refers to observational information gathered directly (in the field) from a natural resource via specific protocols and organized for analysis, summary or reporting. Data may be in either analog or digital form (generally stored either on paper or a variety of computer-compatible media), though the latter is encouraged where ever feasible. Data may exist in several states (conditions) including 'raw', 'validated' and 'analyzed'. 'Analyzed' data includes 'reported' or 'summarized' data and may represent 'information' as a final form of the data from which decisions or conclusions may be made. Ultimately, data are intended to contribute to the knowledge and decisions regarding the conditions, processes, and changes within the ecosystem.

Data set can best be considered a convenient grouping of data, or individual observations, such that the summary of the information will be meaningful to prospective users.

Entity relationship diagrams (ERDs) are high-level data models that are useful in developing conceptual designs for databases. Creation of an ER diagram, which is one of the first steps in designing a database, helps the designer(s) to understand and to specify the desired components of the database and the relationships among those components. An ER model is a diagram containing entities or "items", relationships among them, and attributes of the entities and the relationships.

Indicators are a subset of monitoring attributes that are particularly information-rich in the sense that their values are somehow indicative of the quality, health, or integrity of the larger ecological system to which they belong. Indicators are a selected subset of the physical, chemical, and biological elements and processes of natural systems that are selected to represent the overall health or condition of the system.

Inventories are "an extensive point-in-time effort to determine location or condition of a resource, including the presence, class, distribution, and status of plants, animals, and abiotic components such as water, soils, landforms, and climate. Inventories contribute to a statement of park resources, which is best described in relation to a standard condition such as the natural or unimpaired state. Inventories may involve both the compilation of existing information and the acquisition of new information. They may be relative to

either a particular point in space (synoptic) or time (temporal).” (Source: http://www.nature.nps.gov/im/monitor/index.htm)

Legacy Data are data collected before the GLKN Data Management Plan was implemented. Generally, this refers to data collected in the park units that the GLKN deems important to the monitoring mission.

Logical data model (LDM) is an abstract representation of a set of data entities and their relationship, usually including their key attributes. It may leave out non-key attributes or use abstractions of actual types, and there may be physical tables that are represented in the logical model by relationships or as aggregates within a larger entity, depending on the notation used and the level of abstraction. The logical data model is intended to facilitate analysis of the function of the data design, and is not intended to be a full representation of the physical database. It is typically produced early in system design, and it is frequently a precursor to the physical data model that documents the actual implementation of the database.

Metadata is data about data. It is a document that contains specific and detailed information about a dataset, including who, what where, when, why and how the data was collected, analyzed, or manipulated.

NatureBib is the National Park Service bibliographic database. NatureBib is designed to work with NPSFocus to help staff learn of relevant reports, locate them and where possible obtain digitally.

Natural Resource Database Template (NRDT) is a relational database template that provides a basic underlying relational structure for developing databases. It includes standards for describing location and event information, to which protocol-specific tables and relationships can be added.

NPSFocus is the National Park Service online library designed to provide a medium for literature research and, where possible, digital access to relevant reports.

Ortho-rectification is the process of correcting an image for distortions due to camera lens, sensor non-verticality, and terrain, a process collectively included in photogrammetry.

Orthophoto is a digital image in which the pixels have been corrected to an orthogonal projection, produced by removing errors due to tilt and relief displacement.

Physical data model is a representation of a data design which takes into account the facilities and constraints of a given database management system. In the lifecycle of a project it is typically derived from a logical data model, though it may be reverse-engineered from a given database implementation. A complete physical data model will include all the database components required to create relationships between tables or

achieve performance goals, such as indexes, constraint definitions, linking tables, partitioned tables or clusters.

Protocol, (Monitoring Protocol), refers to the formal documents and sampling processes describing how a vital sign will be monitored; composed of a narrative section, standard operating procedures and supplementary information (databases, reports, tools, hardcopy materials).

Raw Data are data obtained from the environment and that has not been subjected to any quality assurance or control beyond those applied during field work. Typically, raw data constitutes field data sheets but may also include remotely sensed data, data gathered electronically on field computers and photographic imagery.

Relational Database System (RDMS/RDBMS) is a type of database management system (DBMS) that stores data in the form of related tables. Relational databases are powerful because they require few assumptions about how data is related or how it will be extracted from the database. As a result, the same database can be viewed in many different ways. An important feature of relational systems is that a single database can be spread across several tables. This differs from flat-file databases, in which each database is self-contained in a single table.

Sensitive Data are data that is not intended for release to the public, and are not subject to FOIA requests. These data are commonly include information about threatened and endangered species, or archaeological resources, whereby disclosure of such information could result in harm or vandalism.

Standard Operating Procedures (SOPs) are detailed step-by-step instructions for carrying out sampling procedures in monitoring Protocols (the 'How?' of data management planning); dynamic in nature and frequently updated; can reference overall data management plan and standards.

STORET/NPSTORET is a database application maintained by the U.S. Environmental Protection Agency that contains raw biological, chemical and physical data on surface and ground water collected by federal, state and local agencies, Indian Tribes, volunteer groups, academics, and others. All 50 states, and jurisdictions of the U.S. are represented in these systems. NPSTORET is the NPS version of this database used for data entry by NPS staff. Data from NPSTORET are subsequently transferred to STORET.

Structured Query Language (SQL) is a language used to interrogate and process data in a relational database. Originally developed for mainframes, all database systems designed for client/server environments support SQL. SQL commands can be used to interactively work with a database or can be embedded within a programming language to interface to a database. Programming extensions to SQL have turned it into a full-blown database programming language, and all major database management systems (RDBMS s) support the language. There is an ANSI standardized SQL, but most

database management systems (RDBMS s) have some proprietary enhancement, which if used, makes SQL non-standard.

Validated data are data that have been verified according to the standard operating procedure under which the data were gathered (typically the protocol for a given project) and are deemed ready for reporting and/or analysis.

Vital Signs, as used by the National Park Service, are a subset of physical, chemical, and biological elements and processes of park ecosystems that are selected to represent the overall health or condition of park resources, known or hypothesized effects of stressors, or elements that have important human values. The elements and processes that are monitored are a subset of the total suite of natural resources that park managers are directed to preserve "unimpaired for future generations," including water, air, geological resources, plants and animals, and the various ecological, biological, and physical processes that act on those resources. Vital signs may occur at any level of organization including landscape, community, population, or genetic level, and may be compositional (referring to the variety of elements in the system), structural (referring to the organization or pattern of the system), or functional (referring to ecological processes).